JOURNAL OF THEORETICAL BIOLOGY

ACADEMIC PRESS

VOLUME 1, NUMBER 4 OCTOBER 1961

London and New York

Journal of Theoretical Biology

Volume 1, 1961: 12*l*s. 6*d*. ($17.00) Private subscriptions only: 86*s*. ($12.00)
Subscription orders should be sent to Academic Press Inc. (London) Ltd., 17 Old Queen Street, London, S.W.1, except those from Canada, Central and South America, and the U.S.A. which should be sent to Academic Press Inc., 111 Fifth Avenue, New York 3, N.Y.

J. Theoret. Biol. (1961) **1**, 411–414

A Model for the Generation of Self-Sterility Alleles

Sir Ronald Fisher, F.R.S.

Division of Mathematical Statistics, C.S.I.R.O.

(*Received* 17 *April* 1961)

1. Introduction

One of the commonest genetic situations interpretable as a self-sterility mechanism, is that in which a number of alleles exist, presumably at the same locus, with antigenic properties such that the style tissue shall arrest the progress of pollen tubes containing either of the genes of the seed parent. The seed parent is supposed to accept pollen of all other save these two kinds.

As more species showing these peculiarities have come to be studied, two curious facts have emerged.

(*a*) The number of alleles which must exist in most species is remarkably large. Although a population could be maintained with three alleles only, without allowing any illegitimate fertilization, yet it is probable that the number in most natural populations runs to hundreds, and even in the species *Oenothera organensis* which, when discovered, was exceedingly sparse in individuals, yet these contained something like forty alleles (Emerson, 1939 ; Fisher, 1961).

(*b*) The hypothetical mutational process by which a new allele might be produced, has never been observed. Lewis's (1949) extensive studies in self-pollination show no such case, and it would require a mutation-rate less than 10^{-9} in each generation to explain this absence, if new alleles really were being created by mutation.

It appeared to the author more probable that new alleles were synthesized by recombination, and that a new allele formed in this way would not be acceptable in a style secreting antibodies to both parent alleles. In this way the supply of new alleles might be comparatively abundant, without their revealing themselves in the technique of self-pollination.

As the situation proposed is not familiar to geneticists, I shall exemplify its working by a simple hypothetical case, in which, in a population containing only three alleles, a fourth allele is formed by recombination of two of these.

2. The Model Population

Let us imagine a population with three genes A, B and C, and consequently three genotypes BC, AC, AB each of which will accept only one kind of pollen. Each, therefore, produces seed of the two genotypes other than itself. Equilibrium in which each genotype is equally frequent will be approached rapidly, for, as is easily seen, if u, v and w are the three frequencies, the compounds

$$2u—v—w$$
$$2v—u—w$$
$$2w—u—v$$

will each be diminished in the next generation with change of sign, so that in fact it is multiplied by $(-\frac{1}{2})$. Disturbances due to random sampling, often called "drift", are not only trifling in magnitude in a population of no more than 100, but reverse the sign of their effect in each succeeding generation, as they rapidly die away.

Suppose now that the heterozygous genotype AB produces by recombination a gamete carrying a new allele X. The new allele is at no disadvantage in an ovule, and is capable of being fertilized by A, B or C pollen. In a pollen grain, however, it will be effectual on stigmas of genotypes AC, BC. In the male gamete, therefore, it has a 2 : 1 advantage, so that as it appears equally in pollen and in ovule, the ratio of its increase is 3/2 in every generation. In a large population the logarithm (to the base 10) of frequency increases by

$$0{\cdot}17609$$

in each generation ; Table 1 illustrates the effect of such an event occurring in a large inter-breeding population.

TABLE 1

Generation from crossover	Expected number of loci occupied by new alleles
5	7·594
10	57·66
15	437·9
20	3,325
25	25,250

It should be remarked that these expectations are not influenced by the fact that a single occurrence may die out in the first few generations. The probability of this can be easily calculated from the relation

$$\frac{\log_e p^{-1}}{1-p} = 1{\cdot}5,$$

showing that the probability of extinction is in fact

$$p = 0{\cdot}41718835,$$

or not far from 3/7. If the crossover survives, therefore, the expected number of descendants in the table above should be multiplied by about 7/4. The figures in the table are the expectations when it is not known whether the new allele will survive or not.

3. Equilibrium in a Finite Population

The loci containing the new allele will be distributed among the new genotypes AX, BX, CX. As long as these constitute a negligible proportion of the population of each generation, so long will the geometric increase continue.

In principle, the constitution of the population in statistical equilibrium may be ascertained by equating the proportion of the six genotypes to those of their products by random fertilization. In more complicated cases the degree of the equations to be solved may be very high. In the case immediately under discussion, they are of the fourth degree, but the relevant root is rational, and the solution rather simple.

Table 2 enables the correctness of the solution to be checked and exhibits the details of a system of a kind not hitherto discussed.

TABLE 2

Table of reproduction in each generation

Genotype of seed-parent	AB	AC	BC	CX	AX	BX	Total
Relative frequencies	42	52	52	36	26	26	234
Acceptable pollen	C only	B,X	A,X	A,B	B,C	A,C	
Pollen ratio		15 : 11	15 : 11	1 : 1	6 : 7	6 : 7	
Seed genotypes							
AB	—	15	15	—	6	6	42
AC	21	—	15	9	7	—	52
BC	21	15	—	9	—	7	52
CX	—	11	11	—	7	7	36
AX	—	11	—	9	—	6	26
BX	—	—	11	9	6	—	26

The proportions of the four genes in the population are:

A	B	C	X	Total
30	30	35	22	117

and the four types of pollen should occur in the same ratios. However, on any particular seed parent, the relevant proportions are those of eligible pollen only. Even if C pollen were not the most abundant, it would have to account for all seeds born on AB plants.

An observable feature in such a population, if uncomplicated by later recombinations, is that one of the six genotypes, AB, gives seed progenies of only two classes, whereas all others are expected to give four. The source of the new allele is, therefore, recognizable in future generations. In a large outcrossing population with only 3 alleles, grown from seed year after year, at first every plant yields seed of only two genotypes. The first sign of a new allele will probably be the appearance in some seed progenies of a rare third class, fertile to the pollen of both sister types, but with pollen acceptable to only one of them.

In a population of only three genes, grown year after year by open pollination, the initial condition is that progenies from any seed-parent consist of only two genotypes, mutually non-fertile. The first sign of the appearance of a new allele would probably be that a few seed progenies would contain a new rare genotype, accepting pollen from both classes of its sibs, but with pollen acceptable to perhaps only one of them. From such an exceptional plant the hypothetical system, if realized, could be verified in detail.

A short strip of chromosome with ten or twenty antigenically active points, not unlike some of the blood-group loci in man, could provide a thousand or more distinguishable combinations capable of acting as self-sterility alleles. The feature that antibodies effective against two alleles should also be effective against recombinations of them would perhaps show most clearly in such a three allele population as has been described.

REFERENCES

EMERSON, S. (1939). A preliminary survey of the *Oenothera organensis* population. *Genetics*, **24**, 524.

FISHER, R. A. (1930). " The Genetical Theory of Natural Selection ". Oxford University Press; reprinted in Dover Publications.

FISHER, R. A. (1961). Possible differentiation in the wild population of *Oenothera organensis*. *Aust. J. biol. Sci.* **14**, 76.

LEWIS, D. (1949). Structure of the incompatibility gene. *Heredity*, **3**, 339.

J. Theoret. Biol. (1961) 1, 415–430

Functional Analysis of Chemical Systems *in vivo* using a Logical Circuit Equivalent

Motoyosi Sugita

Hitotsubashi University, Tokyo, Japan

(*Received* 9 *January* 1961)

A functional analysis of chemical systems *in vivo* is attempted, using a digital model. It is based upon a hydrodynamic model of an enzymic reaction, in which the reaction is represented as flow under a pressure difference from one tank to another through a control valve. In this model the height of fluid in a tank corresponds to the concentration of chemical substance (or metabolite), and the control valve corresponds to enzyme. The control valve is assumed to be either open or closed, and only the steady state is considered. Binary or two valued logic is used for expressing the existence, or non-existence, of a metabolite. A corresponding logical circuit is then presented. In this way the concepts of information and feedback in chemical systems can be clarified. The presentation is based upon modern thermodynamics, and particularly on the "variable throttling factor" theory (cf. Prigogine's "coefficient d'entraînement"), which is used to take into account the fact that the "throttling factors" are variable parameters.

1. Introduction

Some types of chemical reaction system *in vivo* have *physiological* functions. For the understanding of such functions the author of this paper has suggested (1961a) the equivalence of a certain chemical system to a logical circuit (assuming the *binary* or *two valued logic*) involving the existence (1) or non-existence (0) of chemical substances and the *on-off control* of the rate of enzymatic reactions. In this paper such a correspondence is considered in detail, and the application of the majority principle, which will be described in Section 3, is suggested.

We shall consider the following enzymatic reaction

$$S \xrightarrow{E} P \quad \text{or} \quad S \xrightarrow{X} P, \tag{1.1}$$

where X denotes the complex of the enzyme E and low molecular substances. Figure 1 shows a model of this relation (1.1), where the action of the enzyme E or the complex X and the quantity of the product P are represented by the control valve and the height of the tank respectively.

Let us consider, for simplicity, a chemical system that is structureless

in space and still has a certain logical function. As was assumed above, the on-off control of the valve is taken into consideration. We shall take only the *steady state* into account and neglect the transient state for simplicity. The exit of fluid from the right of the tank of Fig. 1 shows the consecutive reaction process of P to another product. Then P = 0, when the valve is *closed*, while P will attain a certain height in the steady state

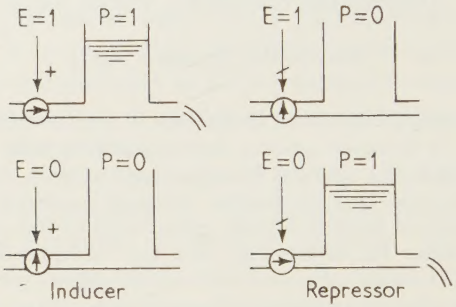


FIG. 1. Fluid model.

when the valve is *open*. Thus, the accumulation of P is neglected. We shall assign 1 to the state of existence of chemical substances, S, P, E or X, P = 1 for instance.

We shall call X *activator*†, if P = 1 when X = 1, but *inhibitor*†, if P = 0 when X = 1.

The *digital approach* of this kind is but a crude model of the reactions *in vivo*; however, we can obtain a method of analysis by such a simplification and thereby see the important feature of these chemical reaction systems, in which the rate of reaction is catalyzed by some sort of products of these reactions. In such systems the flux of the controlled reaction corresponds to the *signal current*, which conversely controls some other valves of the fluid model or the switches of a certain equivalent circuit. From this equivalent circuit we may also get the principles for reconstruction of the living, functioning organism using the components obtained from living things, just as in the reconstruction of a dismantled watch.

According to the modern theory of irreversible processes (Prigogine, 1947; de Groot, 1951) the phenomenological relation of the type

$$\text{flux} = \text{throttle factor} \times \text{intensive factor} \qquad (1.2)$$

† In some cases and figures the term inducer is used for positive catalyzer and repressor for negative one.

is known where, on the one hand, the intensive factor is expressed by the thermodynamic quantities such as chemical potential, and, on the other, the throttle factor is a quantity such as reaction rate constant, diffusion constant, electric conductivity, etc., and determined empirically. Ordinarily these factors are constant parameters.

Catalytic action corresponds to control of the throttle factor. A catalyst cannot catalyze those reactions which are thermodynamically impossible: the possibility of catalysis is numerically expressed by the intensive factor of (1.2). An enzyme may catalyze only those reactions *in vivo*, which are possible thermodynamically and have the *field of the intensive factor* along the *reaction coordinate*. The intensive factor corresponds to the difference of water head of the fluid model. The author has been studying thermodynamics of transient phenomena (Sugita, 1950, 1951, 1954) and has proposed the *flexible throttle theory* (Sugita, 1958), in which the throttle factor is variable instead of constant parameters. Hydrodynamical models having flexible throttling were considered formerly by Franck (1956), who used such a model to explain the excitation of nerve cells. We intend here, however, to use the model for biochemical reactions.

If the product P of $S \xrightarrow{x} P$ controls the throttle factor of another reaction $S' \xrightarrow{P} P'$, then the flux of $S \longrightarrow P$ corresponds to the signal current as was considered above and there is a correlation between the two reactions. Correlation of this type is *indirect*, because only the throttle factor of $S' \longrightarrow P'$ is flexible in this case, and the change of the intensive factor of this reaction is neglected here. There may be no direct interaction between the fluid of two vessels of Fig. 2 of Section 2, for instance. This mode of control is similar to control by a signal current, which acts upon a switch or variable resistance, while the potential of electric source (intensive factor) is kept constant. The author called such an indirect relationship the *informational correlation* (Sugita, 1958), because the energy (or the driving force) for the response comes from "another system" such as a battery or metabolic system, and the input serves only as a signal or a certain pattern of information, which controls indirectly the mobilization of energy of another system.

In this report we shall compare the behaviour of the system of controlled chemical reactions having informational correlation with that of a logical circuit, a *switching circuit* for instance.

2. Logical Equation corresponding to a Controlled Chemical Reaction

Let us consider the correspondence considered above in detail and try to use a logical equation for the expression of the function or the behaviour

of this system. As was considered in Section 1, P = 0 after a short time when the reaction

$$S \xrightarrow{X} P \qquad \text{or} \qquad S \xrightarrow[\downarrow +]{X} P \tag{2.1}$$

has been cut off, where $\xrightarrow{+}$ denotes the flux of information controlling the throttle from S to P and + shows that the enzyme action is inducing. The transient state during the short interval between the cut-off of the flux of (2.1) and the terminal state of P, which is assigned by 0, is neglected. Only the existence (P = 1) or the non-existence (P = 0) is considered here.

In the case of *inhibitor* we use the following notation

$$S \xrightarrow{\overline{X}} P \qquad \text{or} \qquad S \xrightarrow[\downarrow]{X} P \tag{2.2}$$

and

$$X \dashv\!\rightarrow \overline{X},$$

where $\overline{X}$ is the *negation* of X, i.e. $\overline{X} = 0$ when $X = 1$ (existence) and $\overline{X} = 1$ when $X = 0$ (non-existence), and $\dashv\!\rightarrow$ denotes the operation of *Not*. For this purpose a sign changer may be used in the electronic circuit. A biochemical example of the corresponding operation is given later (2.4). In the case of inhibitor action the negation is inherent and there may be no need of the special operation of *Not*. It is very important to differentiate the *flux of information* like $\xrightarrow{+}$ and $\dashv\!\rightarrow$ from that of matter or energy, though the flux of information is associated with that of rate processes just like the signal of electrical communication. If we confuse the flux of information with that of chemical processes or confuse the feedback of information with the cycle or circulation of chemical substances, the application of the concepts of informational technology may be unsuccessful (Sugita, 1960). Therefore we assign thin arrow to the flux of information to differentiate it from the thick arrow of the flux of matter.

If X is an *activator* and X = 1, the steady state of P will attain a certain value, which is assigned by 1. Therefore, we can get the following logical equations corresponding to (2.1) or (2.2)

$$P = X \text{ for activator and } P = \overline{X} \text{ for inhibitor.} \tag{2.3}$$

Therefore, the existence of X(X = 1) corresponds to the existence of P(P = 1) in the case of *activator* and non-existence of P(P = 0) in the case of *inhibitor*.

The following system of chemical reactions has the function of *Not* (see Fig. 2a).

$$A \xrightarrow{\overline{X}} B,\ P \xrightarrow{\overline{B}} Q \qquad \text{or} \qquad \begin{array}{l} \overset{X}{\downarrow +} \\ A \longrightarrow B \\ \quad \downarrow - \\ P \longrightarrow Q \end{array} \tag{2.4}$$

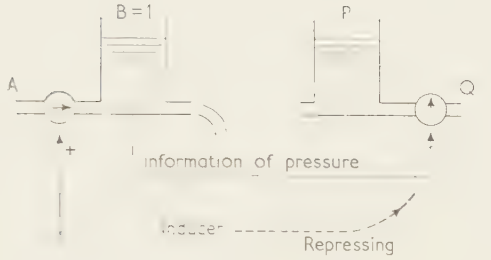


FIG. 2a. Negation.

Then we have

$$Q = \overline{X} \qquad \text{or} \qquad X \dashrightarrow Q = \overline{X}$$

Meanwhile, in the following case

$$A \xrightarrow{\overline{X}} B,\ P \xrightarrow{\overline{B}} Q, \qquad \text{or} \qquad \begin{array}{l} \overset{X}{\downarrow -} \\ A \longrightarrow B \\ \quad \downarrow - \\ P \longrightarrow Q \end{array} \tag{2.5}$$

we can see that

$$Q = X,$$

because $Q = \overline{B}$ and $B = \overline{X}$, therefore,

$$Q = \overline{\overline{X}} = X, \tag{2.6}$$

i.e. the negation of negation of X is X itself (see Fig. 2b), e.g. in biochemistry the repressor of repressor is the inducer. The enzyme induction of

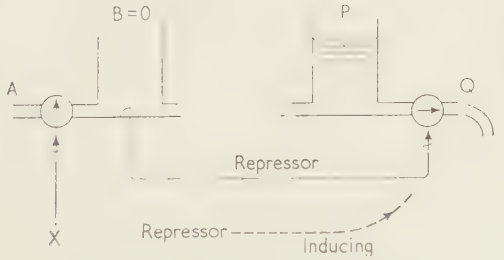


FIG. 2b. Negation of negation.

the type (2.5) is indirect. However, we shall abstract the difference of the *direct* and the *indirect* types of enzyme induction from the logical point of view. According to this viewpoint the following system of reactions

$$A \xrightarrow{X} B,\ M \xrightarrow{\bar{B}} N,\ P \xrightarrow{\bar{N}} Q,\ S \xrightarrow{Q} T,\ \ldots\ldots, \tag{2.7}$$

where B, N, and Q are metabolites (see Fig. 3), can be expressed simply by

$$Q = X,\ S \xrightarrow{X} T, \qquad \text{or} \qquad T = X. \tag{2.7'}$$

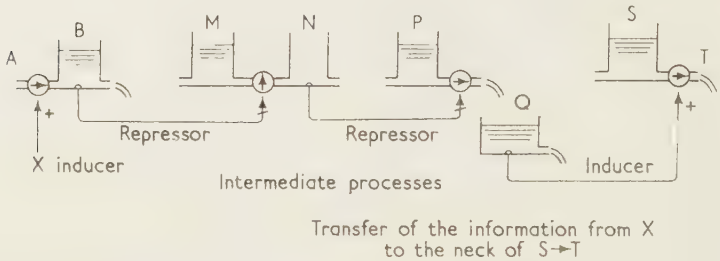


Fig. 3. Indirect action of information.

The effect of negation of even number, $\bar{B}$ and $\bar{N}$, is cancelled out. From the viewpoint of pure logic X may control directly the last valve of T and the intermediate processes may be only the devices of the information transfer.

In this simplification the details of the molecular processes (molecular approach) also are extracted and only the analysis of logical nature (logical approach) is performed. Therefore, we have to imagine a very complicated network of chemical reactions behind the simplified reaction system or logical formulation (see Section **3**).

The sequence of (2.7) correlated by information and control is called *hierarchy*† by the author (Sugita, 1958). The lower reaction, S ⟶ T is controlled by the upper reaction of this hierarchy, A ⟶ B for instance. The informational correlation of this type is to *feed forward* or *ahead*. If the upper reaction of a hierarchy, A ⟶ B for instance, is controlled by the information from the lower reaction like S ⟶ T, such a type of informational correlation is called *feedback*. The genetic control of reactions *in vivo* is usually feed forward, while we can see feedback, if the genetic information is influenced by the lower reactions of hierarchy.

There are two circuit elements besides *Not*, which are at least necessary

† i.e. informational correlation having ranks such as forward and backward or, in other words, not the simple relations of each other on the same level.

for the construction of logical circuit, i.e. *And Gate* and *Or Gate* (*Buffer*).

And Gate is rather easy to consider. Let us take the following consecutive reactions

$$A \xrightarrow{X} B \xrightarrow{Y} C \tag{2.8}$$

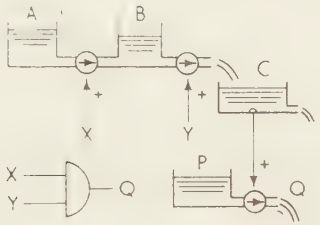


FIG. 4a. Logical product.

into consideration (see Fig. 4a). Assuming binary logic and on-off control of X or Y, we can see that the equation of *logical product*

$$C = XY \tag{2.9}$$

shows the logical nature of the controlled reaction (2.8).

A more complex concept is a chemical system having the character of *Or Gate*, the reaction system of which is in sequence or *consecutive* instead of in *parallel*. (The parallel one corresponds to the conversion of a by-product of a metabolic pathway to the main route, for instance. Consideration of a sequential reaction system having the nature of *Or Gate* was insisted by Sibatani in a letter to the writer.) The author has considered the following scheme (see Fig. 4b)

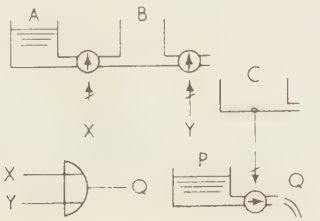


FIG. 4b. Logical sum.

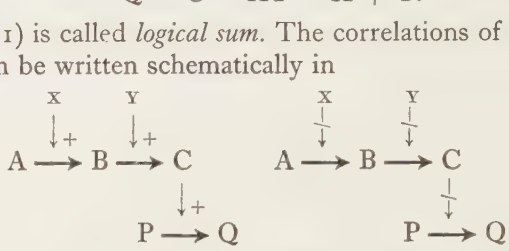
$$A \xrightarrow{\bar{X}} B \xrightarrow{\bar{Y}} C, \quad P \xrightarrow{\bar{C}} Q \tag{2.10}$$

Then we have $C = \bar{X}\bar{Y}$ and $Q = \bar{C}$

From the formula of symbolic logic we have

$$Q = \overline{C} = \overline{\overline{X}\overline{Y}} = X + Y. \tag{2.11}$$

Equation (2.11) is called *logical sum*. The correlations of the reaction (2.8) and (2.10) can be written schematically in



(2.12)

respectively. Examples of more complex reaction systems are given by Fig. 5.

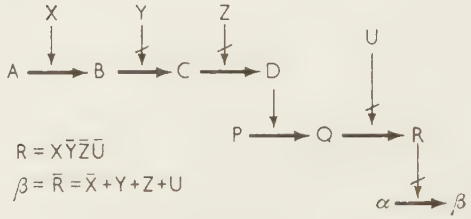


FIG. 5.

3. Logical Equation of a Network of Chemical Reactions and the Application of the Majority Principle of Logical Circuits

Let us consider the following network of chemical reactions

$$\left.\begin{array}{l} \cdots\cdots S''_1 \xrightarrow{X''_1} S'_1 \xrightarrow{X'_1} S_1 \xrightarrow{X_1} P_1 \xrightarrow{Y_1} P'_1 \xrightarrow{Y'_1} P''_1 \cdots\cdots \\ \cdots\cdots S''_2 \xrightarrow{X''_2} S'_2 \xrightarrow{X'_2} S_2 \xrightarrow{X_2} P_2 \xrightarrow{Y_2} P'_2 \xrightarrow{Y'_2} P''_2 \cdots\cdots \\ \cdots\cdots\cdots\cdots \quad \cdots\cdots\cdots \quad \cdots\cdots\cdots \end{array}\right\} \tag{3.1}$$

and assume that a certain step, say $S_i \longrightarrow P_i$ ($i = 1, 2, \ldots\ldots$), of each chain of reactions, i, is controlled by some metabolites of this system of reactions constituting the network (see Fig. 6), where $S_j^{(n)}$ ($n \geqq 0$) and $P_k^{(m)}$ ($m \geqq 0$) are these metabolites.

Let us assume here that X_i is the logical function of these metabolites, i.e.

$$X_i = f_i(\ldots S_j^{(n)}, \ldots\ldots, P_k^{(m)}, \ldots\ldots). \tag{3.2}$$

If $j = i$, then the effect of $S_i^{(n)}$ is *feed forward* and if $k = i$, then the effect of $P_i^{(m)}$ is *feedback*, where $S_i^{(1)} = S'_i$, $S_i^{(0)} = S_i$, $P_i^{(0)} = P_i$, and $P_i^{(2)} = P''_i$.

If j, k $<$ i, i.e. the reaction step j or k is the upper one of the hierarchy, then their effect on X_i is feed forward, but feedback if j, k $>$ i.

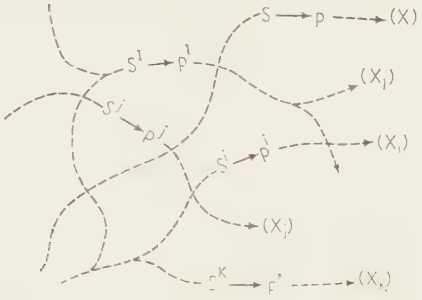


Fig. 6.

Therefore, we can imagine the network or the informational correlation besides the network of chemical reactions. The informational correlation may be feedback as well as feed forward. The consideration given in Section **2** is a simplification of the scheme here considered, so that we have to visualize complicated reaction systems behind simplified schemes such as that of Section **2**.

In general the functional form of (3.2) need not be the logical function of binary logic concerning with X_i, $S_j^{(n)}$, $P_k^{(m)}$. It may be the ordinary function of numerical values of $S_j^{(n)}$ or $P_k^{(m)}$ in the analogue approach. Such an *analogue analysis* may be greatly developed and extended by using an analogue or digital computer. In this report, however, the author will confine himself to the *digital approach* relying upon binary logic and considering on-off control.

$S_j^{(n)}$ or $P_k^{(m)}$ has the value 1 or 0 and their effect is either inducing or repressing.

Although such an approximation is adopted, the analysis may be greatly complicated; so that the application of majority principle is suggested. Such an application was first tried by McCulloch & Pitts (1943) in their logical analysis of nerve activity and then by Muroga (1958) in his study of parametron† circuits.

For simplicity, let us consider a logical function

$$X = f(a_1, a_2, a_3, a_4, a_5), \tag{3.3}$$

where a_1, a_2, . . . are the input to a circuit element subject to the majority

† Parametron was invented by E. Goto of Tokyo University. The computers using parametron are made in Japan, for instance the PC 1 of Tokyo University or the Musasino 1 of the Laboratory of Electrical Communication. The latter was constructed by Muroga.

principle. An example of such an element can be shown by Fig. 7, where B and G denote *Or* Gate and *And* Gate respectively and their outputs

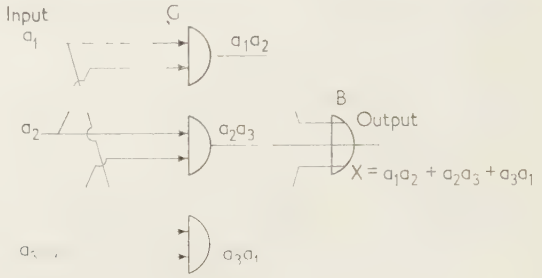


FIG. 7. Majority principle.

are the logical sum and product respectively. The logical equation corresponding to this sort of circuit (inputs are 5 instead of 3) is

$$X = a_1a_2a_3 + a_1a_2a_4 + a_1a_2a_5 + a_1a_3a_4 + a_1a_3a_5 + a_1a_4a_5 + a_2a_3a_4 + a_2a_3a_5 + a_2a_4a_5 + a_3a_4a_5 \qquad (3.3')$$

If the number of input a_k having the value 1 is larger than that of 0 value, then $X = 1$ and, if smaller, $X = 0$. This is the *majority principle*. The application of this principle is suggestive, because a throttle may be repressed by some metabolites and induced by some other metabolites. This is only a trial but, however, suggestive, because the analysis of the function (Elsasser, 1958) of chemical systems is prohibitively complicated, if we do not use such a simplification.

Here we have considered at first the simple chains of reactions, which are not branched as well as not cyclic. We also assumed only one step to a chain which is controlled by the network of information circuit. Then the total system of chemical reactions is regulated by the logical circuit corresponding to the informational correlation. The logical circuit composed of such steps of reactions like $S_i \xrightarrow{X_1} P_i$ corresponds to the computer of a process plant and the processes *in vivo* are similar to those of a chemical plant having computer in the production line. Such a process control can be seen not only at the *cellular level* but at the level of enzymatic reactions, including the formation of adaptive enzyme.

In practice the determination of such steps of reaction or their correlation hidden in the complicated biochemical system may be difficult. We know only a few examples† of such steps and correlations. However, if we

† Transfer of genetic information, for instance.

proceed to such an analysis, we may use symbolic logic as well as *automaton* theory, which is the theory of the switching circuit having time delay.

Complicated systems having cycles or branched chains and many steps that are controlled by the informational correlation will be discussed in future. The *analogue approach* also will be taken into account.

4. Chemical Equivalent of Flip-Flop, Logical Circuit of Storage

Concerning the function of *Storage* we can imagine the system of chemical reactions having this function (Sugita, 1961a). For instance, let us translate the logical circuit of flip-flop (see Fig. 8) to a system

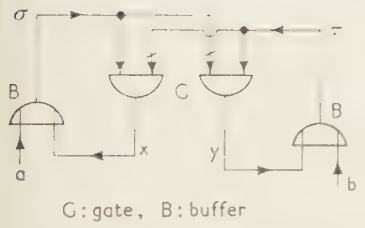


FIG. 8. Logical circuit of flip-flop.

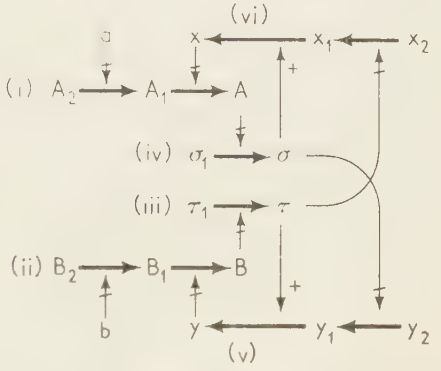


FIG. 9. Chemical equivalence of flip-flop.

of controlled chemical reactions. For this purpose let us consider the following quasi-chemical reactions (see Fig. 9)

$$\left.\begin{array}{l} \text{(i)}\ A_2 \xrightarrow{\bar{a}} A_1 \xrightarrow{\bar{x}} A,\ \text{(iv)}\ \sigma_1 \xrightarrow{\bar{A}} \sigma,\ \text{(vi)}\ x_2 \xrightarrow{\bar{\tau}} x_1 \xrightarrow{\sigma} x, \\ \text{(ii)}\ B_2 \xrightarrow{\bar{b}} B_1 \xrightarrow{\bar{y}} B,\ \text{(iii)}\ \tau_1 \xrightarrow{\bar{B}} \tau,\ \text{(v)}\ y_2 \xrightarrow{\bar{\sigma}} y_1 \xrightarrow{\tau} y, \end{array}\right\} \quad (4.1)$$

where a, x, and so on are catalyzer as well as the reaction products. The logical equations are *in common* to both systems, electronic circuit as well as chemical system of (4.1), and are given by

$$\sigma = a + x,\ x = \bar{\sigma}\tau;\ \tau = b + y,\ y = \bar{\sigma}\tau. \quad (4.2)$$

The analogy of flip-flop and the reaction *in vivo* may be rather crude. However, many types of analogous phenomena are known, which involve switching the course of reaction systems *in vivo*, e.g. mitosis and immunological memory. Assume, for example, that the reactions (ii), (iv), and (vi) of (4.1) are in action and (i), (iii), and (v) are stopped in the resting stage

of a cell, and also that a metabolite b is formed by a certain reaction of a certain chain pertaining to (ii), (iv), or (vi) at the beginning of mitosis; then (ii), (iv), and (vi) are stopped and (i), (iii), and (v) are activated. Now let us assume that a metabolite a is formed by a reaction of a certain chain pertaining to (i), (iii), or (v), then (i), (iii), and (v) are stopped and (ii), (iv), and (vi) are again activated in the daughter cells.

In this reaction system, a and b are the exogenous as well as the endogenous information (or intracellular hormone). The switching of the reactions of (vi), (i), and (iv) and also those of (iii), (ii), and (v) is clear. Here we can see the circulation of information.

In the case of immunological memory the logical circuit or the chemical reaction system may be simpler, as is shown by (4.3)

$$\left.\begin{array}{ll} \text{(i) } A_1 \xrightarrow{\bar{x}} A, & \text{(iv) } x_2 \xrightarrow{\bar{\omega}} x_1 \xrightarrow{\bar{A}} x, \\ \text{(ii) } B_2 \xrightarrow{\bar{b}} B_1 \xrightarrow{\bar{y}} B, & \text{(iii) } \omega_1 \xrightarrow{\bar{B}} \omega, \quad \text{(v) } y_1 \xrightarrow{\omega} y, \end{array}\right\} \tag{4.3}$$

or by Fig. 10a, in which in the normal state (ii) and (iv) are assumed to be in

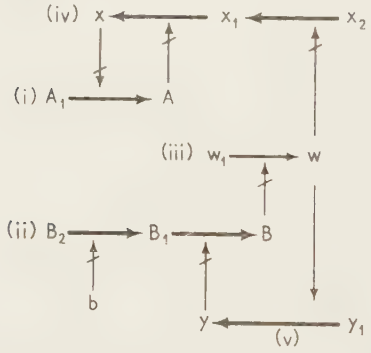


Fig. 10a. Chemical memory.

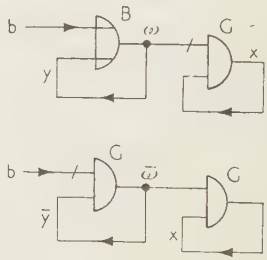


Fig. 10b. Logical circuit of chemical memory.

action and (i), (iii), and (v) are stopped, although these latter reactions are thermodynamically possible, i.e. the intensive factor of these frozen reactions is not zero. If b is introduced (an exogenous factor instead of endogenous in the previous case), then (ii) and (iv) are stopped and (i), (iii) and (v) start into action. In this case the extinction of the reactions (iv) and (i) or (iii), (ii) and (v) is important for the function of *Storage*. The restoration of the initial reaction systems, which were converted by b, is not necessarily wanted.

How to realize the system of reactions (4.1) or (4.3) is neglected here. Only the existence of such a conversion of the mode of chemical reaction

systems is considered here from the logical point of view. There are many possible logical circuits or chemical reaction systems having the same logical function, as is shown by Fig. 10b for instance. They, however, have the same logical equations (4.4), for instance,

$$\omega = b + y,\ \omega = y,\ x = \bar{\omega}\bar{A},\ A = \bar{x} \tag{4.4}$$

in common, so that their function may be expressed by these equations.

The rhythmic reaction of iron in HNO_3 solution may also be interpreted using the analogy of flip-flop. The author, however, is not sure whether the change of ionic permeability of the membrane of nerve cells can be explained in this manner.†

Our translation here considered is merely a trial. We are not yet sufficiently acquainted with the analogy or the correspondence, so that a trial to translate known circuits of electronics having certain functions is done at first. The translation of chemical systems that are well known biochemically, to an equivalent logical circuit is more important and moreover our final aim. Though the details of such chemical systems *in vivo* are not well known yet, the above consideration will suggest a way of proceeding to the experimental analysis or of connecting *biochemistry* with *physiology*.

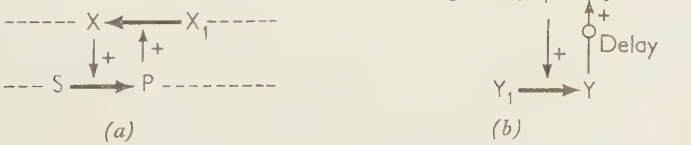
Many other logical circuits having certain functions can be translated into the system of controlled chemical reactions (Sugita, 1961a), details of which are neglected here.

5. Concluding Remarks

Besides the *logical approach* there may be another way of approach which could be called the *molecular approach*. In our simplification molecular details are neglected, so that the difference of the direct and indirect correlation is sometimes extracted and only the logical relation is taken into account.

As to the relation of (3.2), however, we must take the molecular theory

† If we translate Franck's analogue model (1956) to our chemical model, which is digital instead of analogue and in which K is assumed to be maintained by reactions $\ldots X_1 \rightarrow X \ldots$, then we have the following scheme (*a*).



He also presented a model which shows recovery. This model can be translated to that of (*b*), in which the delayed information of Y is considered.

of kinetics into account for the determination of its functional form. We ought to know how the throttle factor of the reaction $S_i \longrightarrow P_i$ is controlled by metabolites, which may form probably the *activated complex* intermediate one between S_i and P_i.

What is then the *information* in chemical systems? From the viewpoint of the digital approach the state, on or off, of the control valve is the information that controls the flux of the reaction rate. From the molecular point of view, the state of the activated complex controls the chemical rate. According to the theory of rate processes their rate is determined by the *free energy of activation*. Therefore, we will take the *entropy part* of this free energy and define the *negative entropy* of information. If this negative entropy is controlled by a certain metabolite, then the flux of the production of this metabolite corresponds to the signal current, which controls the valve of Fig. 1. We may consider here the flux of negative entropy associated with the chemical reaction. Such a consideration seems plausible from the *thermodynamic point of view*† (Sugita, 1961b, in press). In this manner, the idea of informational correlation and of feedback or forward is ensured on the molecular basis.

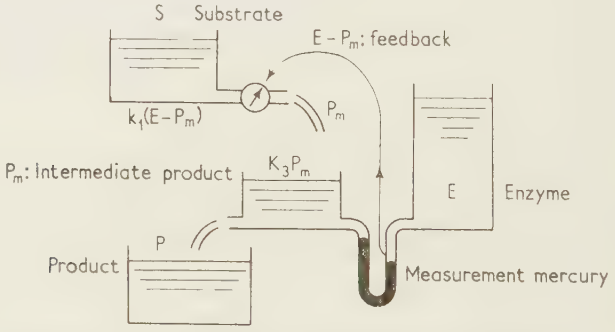


FIG. 11. Direct analogue model of chemical feedback.

† Let the positive value of the entropy of activation, considered per unit volume, be s_a and the vectorial and the scalar fluxes of entropy be $\mathfrak{J}_s$ and $\mathfrak{J}_s^i$ respectively, the former of which is associated with the flux of diffusion and the latter of which is associated with the quasi-chemical reaction i concerning with the complex, then we have

$$\frac{\partial s_a}{\partial t} = -\operatorname{div}\mathfrak{J}_S + \Sigma_i \mathfrak{J}_s^i + \Phi_a, \tag{5.1}$$

where Φ_a is the entropy production pertaining to the rate processes in this volume. If we consider the negative values of s_a, $\mathfrak{J}_s$, and $\mathfrak{J}_s^i$, which are NE_a, $\mathfrak{J}_{NE}$, and $\mathfrak{J}_{NE}^i$ respectively, then we have instead of (5.1)

$$\frac{\partial NE_a}{\partial t} = -\operatorname{div}\mathfrak{J}_{NE} + \Sigma_i \mathfrak{J}_{NE}^i - \Phi_a \tag{5.2}$$

Accordingly we can define the flux of negative entropy. NE_a may be the measure of the molecular order of the intermediate complex.

We consider here only the *digital model* and equivalent logical circuit. The author, however, is considering also analogue models which are equivalent to the processes *in vivo*. There are two types of such analogue

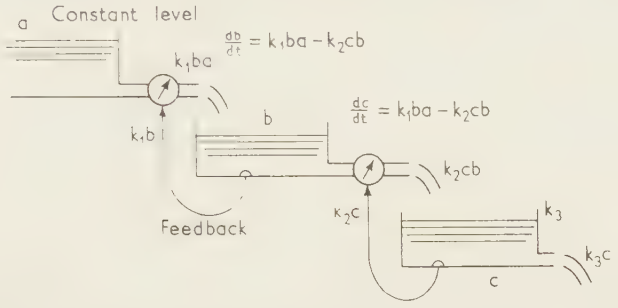


FIG. 12. Direct analogue model of a chemically pulsating system.

models, i.e. *direct* and *indirect*. The former is intuitive and very convenient in considering the nature of feedback of chemical systems. Figure 11 is the direct analogue model which is equivalent to the electronic analogue proposed by Chance (1959) and more convenient than his block diagram of an analogue computer. Figure 12 is the model of the system of autocatalytic chemical reaction proposed by Lotka and discussed by Dembigh *et al.* (1948). The nature of feedback is quite clear in looking at this model. Figure 13 is the model of another system also due to Lotka and Dembigh.

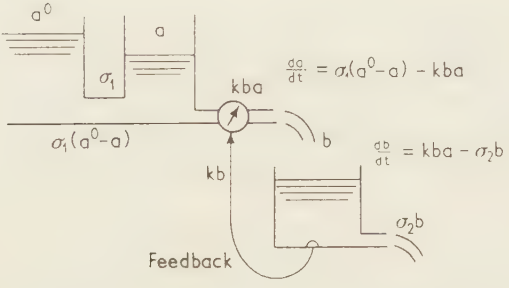


FIG. 13. Direct analogue model of a chemical system having different steady states.

The model of Fig. 13 (Sugita, 1958) is interesting because of its two-valued character† of the steady state. It may serve as a joint device or AD-converter of the analogue and the digital approach.

Dembigh's idea (1952) of the steady state of the chemical system corresponding to Fig. 12 or Fig. 13 is interesting and the author is in accord with his opinion, because the steady state or the minimum

† Franck's model (1956) also has two-valued character of the steady state.

production of entropy may not always be realized in living things. The author also has proposed a theory concerning the maximum rate of consumption of free energy (Sugita, 1950, 1951, 1954). Nevertheless, the utilization of such a physical principle ought to be retracted, because biological consideration is necessary for the ensurance of such a principle.

Although the direct analogue model is intuitive and convenient in our mental picture, the computation or the mathematical analysis in practice may be effectively done by using the indirect analogue which is ordinarily called analogue computer. Dr. Fukuda and the author used this computer for the analysis of isotope kinetics (Fukuda & Sugita, 1960). The computer of this type is used as a sort of general-purpose differential analyser of moderate accuracy, the error of which is from 1% to 0·1%. For an analysis of higher accuracy it is better to use a digital computer, as was done by Chance and his collaborator (Hess & Chance, 1959), or a DDA (digital differential analyzer). For the simulation of the processes *in vivo*, however, the analogue computer may be more convenient, especially for the analysis of the informational correlation in the organism. The author now intends to use a computer of this type as the analogue model of such processes.

Combination of the digital model or switching circuit and the analogue model is also under consideration.

REFERENCES

Chance, B. (1959). In Ciba Foundation Symposium on the Regulation of Cell Metabolism, 355.

De Groot, S. R. (1951). "Thermodynamics of Irreversible Processes". North Holland Publishing Co., Amsterdam.

Dembigh, K. G. (1952). *Trans. Faraday Soc.* **48,** 389.

Dembigh, K. G., Hicks, M., & Page, F. M. (1948). *Trans. Faraday Soc.* **44,** 479.

Elsasser, W. (1958). "The Physical Foundation of Biology". Pergamon Press, London.

Franck, U. F. (1956). "The Progress of Biophysics", Vol. VI, 171. (Butler, J. A. V., Ed.) Pergamon Press, London.

Fukuda, N. & Sugita, M. (1959). *Bull. Kobayasi Inst.* **9,** 1; (1960). *Ibid.* **10,** 1. In Japanese.

Hess, B. & Chance, B. (1959). *Naturwissenschaften*, **46,** 248.

McCulloch, W. S. & Pitts, W. (1943). *Bull. math. Biophys.* **5,** 115.

Muroga, S. (1958). Text of Lecture given at the "Symposium of the Societies of Electrotechnology." Japan Society of Electrical Communication.

Prigogine, I. (1947). "Étude thermodynamique de phénomènes irreversibles". Dunod, Paris; Desoer, Liège.

Sugita, M. (1950). "Thermodynamics of Transient Phenomena". Iwanami, Tokyo. In Japanese.

Sugita, M. (1951). *Ann. Hitotsubashi Ac. Suppl.* **1,** 9.

Sugita, M. (1954). *Bull. Kobayasi Inst.* **4,** 199. In Japanese.

Sugita, M. (1958). *Bull. Kobayasi Inst.* **8,** 159. In Japanese.

Sugita, M. (1960). *Buturi* (Periodical of the Physical Society of Japan, in Japanese), **15,** 528.

Sugita, M. (1961a). *J. phys. Soc. Japan.* **16,** 737.

Sugita, M. (1961b). *Progr. theoret. Phys. Suppl.* In press.

J. Theoret. Biol. (1961) **1**, 431–439

The Avoidance of Over-writing in Self-Organizing Systems†

W. Ross Ashby ‡ and J. Riguet §

Research Laboratory of Electrical Engineering, University of Illinois, Urbana, Illinois, U.S.A.

(*Received 23 January* 1961)

When a whole adaptation or computation must be achieved by stages, so that earlier results must be preserved for use in later stages, the intervening processes must not be allowed to over-write, and destroy, the results found earlier. In today's computers, the programmer looks after this possibility, but when systems become self-organizing, with a random element, over-writing internally is only too apt to occur. The brain has long had to face the difficulty, and developing computers (when more brain-like and self-organizing) will encounter it as a major difficulty. It is a major part of the problem of adaptation to the recurrent situation. We ask, then, how *in general* is over-writing to be avoided in self-organizing systems. What are the general principles that govern the matter?

It is shown here that, quite generally, there are two and only two ways in which over-writing can be avoided (or minimized). Either (i) information in the necessary *quantity* must be processed and utilized, or (ii) the general parameters of design must be so chosen as to reduce the *chance* level to a satisfactory degree; and the second way also demands that an appropriate quantity of information be processed. These deductions are not in themselves surprising, but our aim is to demonstrate that these two methods are the only two, so that the limitation is binding both on the brain and on any constructible computer.

1. Over-writing and Information

The Problem

The problem considered here arises naturally, and indeed inevitably, as soon as a computer (or a living brain) has to undertake the task that is really complex. Then it commonly happens that the task is quite unmanageable if attempted in one piece, but becomes quite manageable if

† Part of work supported by the Office of Naval Research under Contract NR 049–149, at the Burden Neurological Institute, Bristol, England.

‡ Research Laboratory of Electrical Engineering, University of Illinois, Urbana, Illinois, U.S.A.

§ Department of Mathematics, Faculty of Sciences, University of Paris, Orsay, France.

achieved in a sequence of stages. Thus no one can learn mathematics all in one piece, but it is quite manageable if one learns first addition, then subtraction, and so on, each stage being properly completed before the next is tackled.

When the whole task can be (partially) broken up in this way it is "reducible"; and the change to the reduced, or stage-to-stage, form may change the task from the utterly unmanageable to the quite readily achievable. Such a reduction is of the highest importance in really complex adapting (or computing), and becomes of even greater importance as the complexity rises.

The programmer achieves such a reduction when he "sketches out" a program, indicating a sequence of sub-goals, each of which can be worked to with only slight reference to the other stages. The living brain achieves such a reduction when it is able to convert the Grand Problem of "Live long enough to reproduce" into a sequence that is equivalent, such as: First get some air into your lungs; then get some food to eat; then find a secure place to sleep; and so on.

If the sub-tasks should recur, the whole task is then that of the "recurrent situation", discussed by us elsewhere (Ashby, 1960a), and there shown in its great commonness and importance as a major resource in the adaptations of living organisms. It is hardly necessary for us to stress further the great importance of such sequential processes. It is certain that if computers are to equal and surpass human "intelligence" most of their processes must follow this stage-by-stage course.

Suppose, then, that we have such a process, following such a stage-by-stage progression. (Let us exclude the degenerate case in which the "whole" really consists of independent tasks or problems, so that each stage can be carried through without any reference to the events in the other stages.) Call the stages, for clarity of illustration, A, B, C, . . . etc. Then we are considering the case in which stage M, say, requires reference to results obtained earlier, in stage G say, with stages H, I, J, K, and L intervening. Now all these processes . . . G, H, . . . , M are assumed to be occurring in one system, so the value (or values) generated in stage G must persist *un*changed in a system that is to go through the *changes* implied by the processes of stages H through L. Thus, in any system that proceeds by stages and requires reference back, it is essential that in some way the results of earlier stages be protected from the changes implicit in the processes that occur between the results' establishment and their use. (By "over-writing" in the title we meant precisely what would occur if results established earlier were altered by processes occurring later.) How, *in general*, is such over-writing to be prevented?

Surprisingly, the problem has so far received little attention. Lashley

glanced at it thirty years ago, but most brain physiologists have evaded it, though any study of *interaction*-processes in the brain must ultimately encounter it. Why does not later learning wreck the earlier? Plenty of theories exist about how one act of learning affects the brain, but not one, so far as we know, has made any serious attempt to say what will happen when a second, and different (perhaps unrelated) set of learnings is piled on top of the first! Yet the brain does accumulate learning.

The same problem appears urgently in the theory of computers as soon as we consider the computer that is self-programming. How is a self-programming computer, having worked up some results for use later, to avoid over-writing them during the later stages, when the distribution of activities is under its own control?

In this paper we have made no attempt to devise some mechanism that is *sufficient* to prevent over-writing—a host of them are probably possible. Rather we attempt to show what is *necessary*, so that we can put some bounds to what must occur in the brain, and to what must go into the really advanced computer.

SELECTION IS NECESSARY

The first point is fundamental (and so obvious that it can very easily escape notice!). If the intervening processes (H through L above) cause changes in some set of variables, and do *not* cause changes in some other set (which might be called "carriers" for convenience, without other implications), *then the confining of the processes H through L to the complement of the set of carriers is an act of selection.* Thus, in a computer with 1,000 stores, "keeping the process out of stores 000 through 099" implies that at each step the process must be confined to the particular subset of 900 stores.

With this observation we couple the postulate (Ashby, 1960b): *Any system that achieves appropriate selection* (to a degree better than chance) *does so as a consequence of information received.* It calls for small justification, for any system not conforming to it would at once be recognized intuitively as peculiar. It would be broken if any examination candidate gave the appropriate answer before he was told the question, or if a man filled in a correct claim for insurance before the fire broke out! Whenever an exception seems to occur, both the man in the street and the scientist act in the conviction that the postulate must hold, and that if the situation is examined more closely, the final facts will be found to be in accord with it.

INFORMATION IS NECESSARY

"Not over-writing", then, is an act of appropriate selection, and it can be performed in any system only by the processing of information. The

law of requisite variety (Ashby, 1956) is now applicable, and we can say that the information must be processed in a definite minimal *quantity*. The quantity is in fact readily identifiable; it is equal to the difference between the two entropies, or varieties, in:—

(i) the distribution of processes over the variables of the system when chance factors alone control the distribution;
(ii) the distribution obtained when the processes are restricted so that over-writing does not occur.

(As will be noticed below, we assume here that the system works in conditions sufficiently well defined statistically to give a defined sample space.)

The prevention of over-writing is therefore related in a fundamental way to the two questions:—

(i) *How much* information is available for the selection that prevents it?
(ii) *Whence* does it come?

In the elementary cases, these questions are hardly worth notice, for everything is obvious. The programmer who wishes to find $(a + b)^2 + (c + d)^2$, for instance, finds $(a + b)^2$, puts it in a store, and then instinctively avoids using that store while finding $(c + d)^2$. Or if a sub-routine is to be used later, he puts it in a block of stores and uses them for nothing else. Obviously, in this case the amount of information provided is that necessary to guide the processes to the unreserved locations, and the source of it is the programmer himself. When the amount is small he provides it without a thought; when it is large he provides it with some effort.

Does the brain give us an illustration? Unfortunately, so little is known about the facts of its storage, distribution, and over-writing that little can be gained from it to help us. *We* can apply to *it*, however, the logic of mechanism, and we can say with assurance that later processes in it can avoid over-writing only if information, to the necessary quantity, is available to it, and is processed by it. (This information is, of course, precisely that which must travel through the gating control Γ, in the "recurrent situation" of "Design for a Brain", Chapter 10, 1960a.)

Doubtless, in some of the brain's activities, so much has recurred over the generations that natural selection will have made the necessary selection for the avoidance of over-writing. If, for instance, every generation faced problems that involved touch and the arm, and also problems involving smell and the tongue, but not the other combinations, then natural selection would soon evolve a form with two unjoined centres, one for touch-arm problems and one for smell-tongue, so that neither over-wrote the other. Here the selection is performed ultimately by the environment—by the killing agent—which destroys some anatomical arrangements while permitting others.

WHEN INFORMATION IS LACKING

We can now turn to the case of special interest here—that in which the requisite amount of information is not to be had. The case is of outstanding importance for two reasons.

First, this is the case in the brain with respect to all those serial adaptations in which the stages and processes have not been determined by natural selection—those that are peculiar to the individual, and thus comprise most of his adult learning. Such are those that depend on the language spoken around his birthplace, on the conventions of his society, and on his individual economic activities. At least at first, information about how the various stages should be distributed in the brain is simply not available. The case is thus of major importance in our understanding of the normal processes of education.

The case when information is lacking is also of interest because it is the case that inevitably tends to develop as computers become larger and larger, and as the channel through the programmer, of fixed capacity, becomes less and less adequate for the transmission of the necessary quantity of information. Today this source of trouble is small, but a sophisticated theory of the really large self-organizing computer must prepare to meet it.

What can be said about these cases, when the necessary quantity of information cannot be supplied? What has been said shows inescapably that *the corresponding degree of over-writing is inevitable.* The living brain must work subject to this handicap, and so must any type of computer that works by the ordinary relations of cause and effect. This is our first deduction.

It should be appreciated that this deduction does not rest on some appeal to a peculiarity of electronic hardware, or to some feature of neuronic physiology, or even to any fundamental law of physics and matter; all these could be varied freely without in the least affecting the force of the deduction. It rests on the logic of mechanism, used by carefully tracing chains of cause and effect through well-defined processes. Were someone to claim that he had found an exception, we would at once reply that his device must work by pure magic, able to achieve appropriate effects without being given the appropriate causes, and thus of the same nature as the fantastic examples given earlier. Such a claim would be in the same class as one for a new computer that is said to give the answer without waiting for the program-tape to be run in!

It may be concluded, then, that the inevitability of over-writing in the conditions given, is absolute. Brain and computer alike have this fundamental limitation.

2. Minimizing the Amount of Over-writing

It is of the very essence of Part 1's thesis, that, in certain conditions, destruction of previously built-up programmes (or adaptations, or learning) is inevitable. To evade the harshness of this somewhat unpleasant deduction is to miss whatever is of value in that Part. Nevertheless, not everything on the subject has yet been said; and in this second Part we will show how, by variation of the conditions, the harshness may perhaps be mitigated.

As we have seen, without information the amount of over-writing cannot be reduced below the chance level. (By the latter is meant the amount of over-writing that is produced when the chance factors, i.e. those variables other than those involved in the information-processing, vary in accordance with whatever sample space is operative in the system's conditions or circumstances; the sample space is here assumed to be sufficiently well defined.)

The *total* quantity of over-writing thus depends on both the quantity of information available and on the chance level. Part 1 considered the relation between the quantity of over-writing and the quantity of information *with the chance level fixed*; we can now, however, consider variations in the chance level.

VARYING THE CHANCE LEVEL

This aspect has important applications; for variations in this factor can affect the actual (or total) amount of over-writing when the quantity of information cannot be varied, by being fixed at zero perhaps. It is of special interest because the analysis of Part 1 shows that it is the *only* other factor at work. So if over-writing is to be reduced to a level below that made possible by the information supplied, the reduction can be achieved *only* by alteration at the chance level.

If no particular sample space over the chance factors is assumed, full generality is retained but the topic is left wholly arbitrary. In the general case this is appropriate, for exactly how much over-writing occurs in a particular case must depend on the exact details of the program, and process, and the system.

There is one case, however, that deserves a passing glance, for it is always central in probabilistic theories; it is the case in which the various probabilities of the sample space are all equal and in which the various events are statistically independent. Such a case would occur if a digital computer were built so that, if an order specified no address, it used a random address (with all numbers equi-probable, and independent in sequence). In such a machine, with 4,000 stores say, and two processes

each requiring 20 stores, the chance that the two processes, unguided, would use non-over-lapping sets of stores is that the second set,"drawn without replacement", should be all from a particular 3,980 of the 4,000. It is

$$\frac{(3{,}980\ !)^2}{3{,}960\ !\ \ 4{,}000\ !}$$

i.e. about 0·9; and the chance of over-writing occurring is about 0·1. If, however, (still with no information available) the same processes were transferred to a similar machine with 40,000 stores, the chance of over-lap would fall to about 0·01. Thus, with this type of random distribution the amount of over-writing can be reduced as low as we please, simply by making the machine sufficiently large. In general, we can say that changes at the parametric level (i.e. at a design level before the actual flow of selective information occurs) may be able to reduce the amount of over-writing to any desired degree.

What values are to be given to the parameters of design must be determined by the values of the statistical parameters relevant to the sample space. Again a selection of appropriate values (on the parameters of design) is required, and again (by the postulate given earlier and the law of requisite variety) this can be done only by a processing of the relevant information (about the statistical parameters) in sufficient quantity.

It follows that the minimizing of over-writing by general design is possible only so far as the statistical properties of the chance factors are known. To the degree that they are unknown, or unspecified, *to that degree is the designer absolutely unable to prevent over-writing.*

CONSEQUENCES FOR COMPUTERS

The inferences to be drawn are now fairly obvious but are perhaps worth re-statement for clarity.

When computers of a new order of size are available, the super-problems that they will tackle will almost certainly have to be solved by processes that go stage by stage (even in machines with many parts working in parallel), so that early results will have to be stored for use later and kept inviolate through the intervening processes. The extent to which this can be done depends on the degree to which either of two types of information can be used:

(i) Information about storage in detail, so that appropriate stores are selected;

(ii) Information about the sample space (statistical distribution) of chance factors, so that a statistically appropriate design can be selected.

To the degree that these informations, in any particular case, are not available, to that degree must over-writing be accepted.

The chief point of this paper is that it puts an end to extravagant hopes by showing that the avoidance of over-writing has certain fundamental limitations. Projects for developing computers that respect these limitations will be more realistic (and therefore eventually more successful) than those that attempt to be "more brain-like" by being based on an essentially superstitious belief that the brain has some method or technique that transcends the limitations.

CONSEQUENCES FOR BRAIN-PHYSIOLOGISTS

These deductions are likely to be of assistance in clarifying our ideas about the brain, for they say something of what *must* occur in it.

It is obvious enough that the living organism, and especially the human, is often faced with situations that have to be dealt with stage by stage, and in which later stages require reference back to results obtained earlier. It is also obvious that a most valuable power is that of being able to use, at a later date, adaptations learned at an earlier. These are the processes of serial adaptation and of adaptation to the recurrent situation already treated by us elsewhere (Ashby, 1960*a*). There can therefore be no doubt that the problem of over-writing is an important one in the brain.

As was said earlier, natural selection will look after cases that are the same in detail from generation to generation, by developing a suitable mechanism to be inborn. Of more interest at the moment, however, are those problems that are peculiar to the individual. Here Part **2** shows what can *and must* be done. If the avoidance of over-writing is important to generation after generation, and *constant* in being important (though varying in what is not to be over-written) then natural selection can, and and will, force the genetic parameters to values that show in the brain by its being, through the chance factors, as little affected by over-writing as possible.

What can be said next depends on the type of mechanism at work. If it is devoid of a metric (like the digital computer), the change of a single value makes the mechanism different to an arbitrary degree, making perhaps the new activity totally unlike the previous. Such mechanisms will require the chance factors to be so received or acted on that the *probability of over-lap* is minimal. If, however, the mechanism has a natural metric (like the analogue computer, and like the digital computer when it works on the sums of many little contributions), over-writing, even if it occurs always, may nonetheless be tolerable if it never exceeds a certain small amount. Such mechanisms will require the chance factors to be so received that the *fraction* of over-lap (of process and store) be minimal.

These considerations enable us now to say something about how stored (learned) information should be localized in the brain. So far as the information is peculiar to the individual and is not to be over-written, to that degree must it use some form of localization (for the storage) that is (i) random (because of the lack of information about where it should be stored), and (ii) limited to a small fraction of the total storage (to have a high probability of evading destruction). Whether the stores in the used fraction (for a particular adaptation) are anatomically compact or scattered is irrelevant in the present context. It is well known, however, from the works of Pavlov, Lashley, and many others, that localized removals of tissue do not, in general, destroy a particular few reactions or adaptations, but tend to affect almost all, each to a small degree. In fact, when the cybernetic argument is joined to the physiologist's facts, the conclusion is evident that the stores are distributed widely over the anatomical regions, that the distribution may be determined by local factors and by the characteristics of the stimuli involved, but cannot (when the learning is individual) be governed in detail by considerations of over-writing.

This point of view enables us to appreciate the fundamental nature of retroactive inhibition when it occurs in individual learning. It occurs in such learning because the information which alone could prevent its occurrence is not available at the neuronic points at which the distribution is determined. It is fundamental in the sense that if we build advanced computers that *are* self-organizing, we must accept as inevitable that a a certain degree of retroactive inhibition in them will occur. All we can do as designers is to minimize its degree.

REFERENCES

Ashby, W. Ross (1956). "An Introduction to Cybernetics". John Wiley & Sons, New York; Chapman & Hall, London.

Ashby, W. Ross (1960a). "Design for a Brain", 2nd Edition. John Wiley & Sons, New York; Chapman & Hall, London.

Ashby, W. Ross (1960b). *New Scient.* **7**, 746.

J. Theoret. Biol. (1961) **1**, 440–459

Mathematical Analysis of Metabolism using an Analogue Computer: I. Isotope Kinetics of Iodine Metabolism in the Thyroid Gland

NUBUO FUKUDA

National Institute of Radiological Sciences, Tokyo, Japan

AND

MOTOYOSI SUGITA

Hitotsubashi University, Kunitachi, Tokyo, Japan

(*Received* 30 *January* 1961)

1. Applications of analogue computation and the phenomenological approach to isotope kinetics are discussed.
2. A flow chart of iodine metabolism in the thyroid gland is presented, surmising that I_2 is incorporated into globulin in epithelial cells.
3. A set of fundamental simultaneous equations of 17 independent variables describing the dynamical behavior of ^{131}I in the thyroid gland is presented.
4. Most system parameters are estimated from steady state considerations and significant relations between them are derived.
5. The sets of equations are solved by analogue computation.
6. By varying the values of the system parameters, variations in the model system were obtained which suggested resemblance to pathological states in man.

1. Introduction

There are many difficulties in the mathematical study of isotope kinetics. We shall first discuss this problem in general terms and then consider a particular application, the metabolism of ^{131}I in the thyroid gland. A flow chart has been constructed of this aspect of metabolism and from this a set of simultaneous differential equations involving 17 variables has been derived.

Some of the computational difficulties of such problems can be surmounted by the use of an analogue computer. Chance (1955) has been a pioneer in this field. In our work the number of variables is large, so a computer of corresponding capacity is required. Fortunately we have enjoyed the cooperation of Professor Jinbo's laboratory at Meiji University as well as use of their analogue computer.

Initially it is necessary to determine the system parameters of our differential equations. We obtained approximations which we believe to be of the correct order of magnitude from steady state considerations. We then compared the computer output with experimental data and believe that the flow chart shown in Fig. 1, and especially our assumptions concerning the incorporation of I_2 into the tyrosine residues of thyroglobulin, may correspond to reality. We believe that information on the mechanism of normal and pathological metabolism of the thyroid gland can be obtained from curves observed in medical practice. We suggest that the analogue computer is not only a device for computation but also a powerful tool for the simulation of biochemical systems, whose behavior can then be studied by numerical experimentation.

Chemical and other rate processes in the living organism are of tremendous complexity, so that the information with which we start our mathematical deductions is incomplete. However, in this new age of electronic computation a complicated system can be replaced by a "black box" whose output can be tested in a few seconds, so that we can vary both the initial assumptions and constant parameters freely and frequently, and in each case compare the output with the experimental data.† Thus some of the limitations formerly inherent in the study of complex systems now no longer apply. Of course, the computer can do nothing unless provided with determinate values of system parameters. The exact values of these are unknown, but certainly they are positive and lie between zero and infinity. Taking any values in this region initially, we can search for optimal values in the sense of giving an output corresponding best to empirical data.

2. Mathematical Considerations

If mathematics is to express adequately the *in vivo* situation it must be of the same order of complexity as are such processes as active transport, biosynthesis or metabolic turnover. Even in the case of isotope kinetics, where we can use linear analysis, the mathematical difficulties are still very large. However, we can simplify the mathematical relations by considering suitable compartments and grouping variables. A compartment may be localized in space, or it may be a state of chemical combination which is assigned by a certain value of chemical potential‡ (Sugita, 1954).

† It is rarely possible to determine the unknown parameters as the unique solution in mathematical terms by using such a method. However, the idea of the "black box" is stimulating and reduces the effort required to obtain useful knowledge.

‡ From the thermodynamic point of view, chemical potential is defined as

$$\mu = \frac{\partial G}{\partial n},$$

In this paper we consider the biochemical system to be in a steady state, in which the total quantity of any element in a compartment is constant in time, only the percentage of an isotope of that element being variable. Let x_i be the quantity of a radioactive isotope of an element in compartment i. Then the number of independent variables x_i is finite and is equal to the number of interrelating compartments. Variables other than x_1, x_2 . . . can be considered constant parameters in the steady state.

We then obtain the following phenomenological equations:

$$dx_i/dt = \textit{influx-outflow} \tag{1.1}$$

corresponding to the law of conservation of matter (Sugita, 1954). The influx can be written as:

$$\textit{influx} = I_i + \Sigma_j \mathcal{J}_{ji} \tag{1.2}$$

where I_i is the input of that isotope from the external system directly to i, for instance by injection. $I_i = 0$, of course, if there is no direct input to i. $\mathcal{J}_{ji}$ is the flux from a compartment j to i. If i and j are chemically different compartments, then $\mathcal{J}_{ji}$ is the scalar flux associated with the chemical reaction $j \rightarrow i$. If they are different compartments in space, $\mathcal{J}_{ji}$ is the vectorial flux associated with processes such as passive diffusion or active transport.

The outflux is:

$$\textit{outflux} = e_i + \Sigma_k \mathcal{J}_{ik} \tag{1.2'}$$

where e_i is the excretion from i to the external system and $\mathcal{J}_{ik}$ is the flux from i to k.

In the case of a steady state we can use the following linear expressions:

$$\mathcal{J}_{ji} = K_{ji}x_j,\ \mathcal{J}_{ik} = K_{ik}x_i,\ e_i = K_i x_i \tag{1.3}$$

for the quantities $\mathcal{J}_{ji}$, $\mathcal{J}_{ik}$, e_i of isotope kinetics. K_{ji}, K_{ik} and K_i are constant parameters in the steady state. Again, $K_i = 0$ if there is no direct excretion from i; (1.3) corresponds to the phenomenological relations (Sugita, 1954, 1960) of flux and intensive factor in the modern theory of thermodynamics of irreversible processes. Such phenomenological considerations may be very useful in our case, since the molecular processes are very

where G is the free energy of a system, to which the particles considered belong, and n is the mol number or the number of particles. From the viewpoint of statistical mechanics chemical potential is the mean value of a certain intensive quantity through n particles, which can be grouped thermodynamically. If we take, on the other, the grand partition function Z_g and put

$$pV = -kT \ln Z_g,$$

we get

$$n = \frac{\partial pV}{\partial \mu}.$$

In this case n is interpreted as the mean number of particles in the compartment designated by μ, i.e. having the same value and the same functional form of chemical potential.

complicated and a thermodynamic method is required. (Sugita, 1954, 1960).

If we put the expression (1.3) into (1.2) and (1.2′), then

$$influx = I_i + \Sigma_j K_{ji} x_j, \quad outflux = (K_i + \Sigma_k K_{ik}) x_i \tag{1.4}$$

and (1.1) can be transformed to

$$dx_i/dt = \Sigma_j K_{ji} x_j - (K_i + \Sigma_k K_{ik}) x_i + I_i, \tag{1.5}$$
$$(i : 1, 2 \ldots n),$$

where i is the number of compartments; (1.5) are the fundamental differential equations of isotope kinetics (Fukuda & Sugita, 1959, 1960) and K_{ji}, K_{ik}, K_i are the system parameters. If these parameters are known, as occasionally happens, we can determine the behavior of every x_i by solving the simultaneous equations and comparing the results with experiments. However, such cases are of little practical significance for the following reasons:—

(a) The number of compartments is very large, so that even if approximate methods such as grouping variables are adopted, an analytical solution of differential equations with a very large number of independent variables is in practice prohibitively difficult.

(b) It is very difficult to obtain reliable experimental knowledge of the system parameters.

(c) It is hardly possible, in general, to draw up an exact flow chart of isotope behavior in the organism. For instance, it is not known where I_2 is incorporated into the tyrosine residues of the globulin molecule.

Computational difficulties mentioned under (a) may be met by using either an analogue or a digital computer (Chance & Hess, 1959). Difficulties listed under (b) and (c) are interrelated, since the phenomenological constant K_{ji} is qualitatively related to the flow chart. Using the phenomenological approach, we can interconnect every compartment by an arrow designated K_{ji}. Any K_{ji} that is zero can, of course, be omitted from the flow chart; *a priori*, however, we can say nothing as to the magnitude of any parameter from a knowledge of the flow chart alone.

We shall now consider the application of an analogue computer to overcome the difficulties listed above.

3. Applications of an Analogue Computer

Analogue computers are of two types, the direct and indirect. The former is usually referred to as a mechanical or physical model, while the latter is what is called an analogue computer in the general usage of the term. Many analogue computers built in Japan use thermionic valves and the analogue quantity is generally an electrical potential. This type of computer is merely a kind of differential analyser of low accuracy.

The number of independent variables in a system of differential equations that can be handled by such a computer is roughly determined by the number of operational amplifiers in its electronic circuit. The computer at Meiji University designed by Professor Jinbo has a rather large capacity with 32 operational amplifiers, so that differential equations with more than 20 variables can be handled. It can be used both in high speed and low speed applications. The most useful application of such a computer is in searching for the most suitable parameters, inspecting the curves which are given almost instantly. Professor Jinbo's idea is to search for the optimum region using a high speed of operation, then, using the same parameter value change to a slow speed to obtain a more accurate curve. We, however, have only used a slow speed, as we had at that time insufficient experience with biological and medical applications. The use of high speeds for biochemical research will be developed in the near future.

Even at low speeds computed curves can be drawn in a few seconds, but comparisons with experimental data take much more time. Changing the trial values of the system parameters takes a few minutes. A trial plug board is used to change the flow chart and for this also a few minutes are required. Thus we can easily change both our parameters and flow chart arbitrarily in search of optimum values which best fit the experimental data, although not the only one. From this point of view an analogue computer is not only a device for computation but also a tool for numerical experimentation by the method of trial and error.

The difficulties in handling complex systems cited above thus seem partly to disappear, so that the old approach before computer technology developed is no longer a limiting factor. Nevertheless, it is still not easy to search for optimum values by comparing the output of a computer with experiment, and a knowledge of biochemistry is still required. Since, however, we can search 2 or 3 parameters the data need not be complete. A search of more than 4 parameters is still, however, quite laborious.

As an example of our approach, which we develop in detail below, we consider the metabolism of I_2 in the thyroid gland. We assumed that I_2 is incorporated into the tyrosine residues of globulin in the epithelial cells and constructed the flow chart shown in Fig. 1. In view of the fact that monoiodotyrosine, diiodotyrosine, triiodothyronine and thyroxine (M, D, T_3, T_4) cannot be detected in these cells, our assumption may appear rash. However, on testing our flow chart by analogue computation we found that the computed quantities of the residues, M, D, T_3, T_4 in epithelial cells turned out to be very small and their life time very short, the macromolecules which have incorporated I_2 being transported with a large rate constant k^i into the thyroid follicle. Therefore they would be difficult to detect in these cells experimentally. We had surmised this intuitively

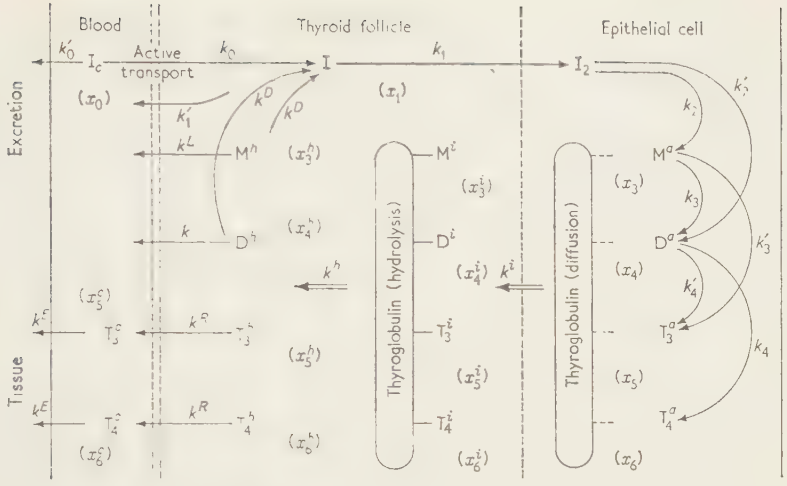


FIG. 1. Flow chart of iodine metabolism in the thyroid gland.

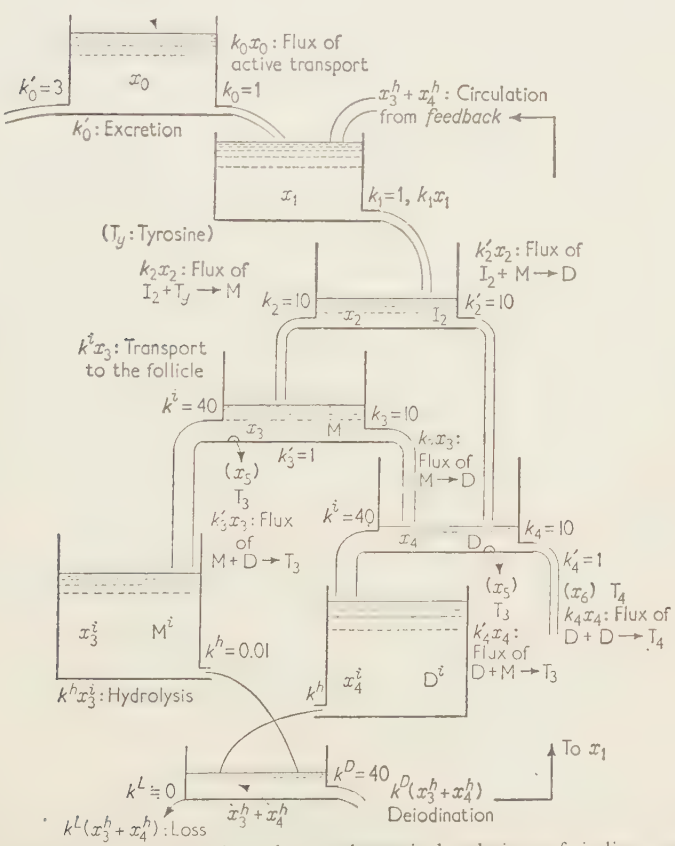


FIG. 2. Fluid model expressing the mathematical relation of iodine metabolism. (Processes after x_5 or x_6 are omitted for simplicity.)

and found our intuition confirmed by computation, provided that the system parameters are adequately determined (see Section 4). An intuitively constructed hydraulic model of the above system is shown in Fig. 2.

While the assumptions we made are supported by trial computations, the rejected assumptions must also be tested. This is the spirit of the method of trial and error. Unfortunately, we have not had enough experience in the use of this type of computer and therefore have experienced many mistakes and failures. For lack of time many things remain to be tested.

4. Analysis of Iodine Metabolism in the Thyroid Gland

We assume the flow chart shown in Fig. 1. Iodide ions, including the radioactive isotope, are trapped and concentrated in the thyroid follicle by active transport. Let the rate constant of trapping be k_0. The I^- ions are then transported into the epithelial cells, oxidized to I_2 ($2I^- \xrightarrow{k_1} I_2$) and I_2 is then incorporated into the tyrosine residues of globulin. The latter then diffuses into the thyroid follicle. Interconversion of iodinated side chains (as $M + D \rightarrow T_3$) is assumed not to occur in the follicle, so we assign the notation i to M, D, etc. On the other hand we assign the notation a to these components in the epithelial cells.

The iodinated side chains of thyroglobulin are hydrolyzed enzymatically in the follicle with the rate constant k^h. This may well be a very slow process. The products M^h, D^h, T_3^h, T_4^h are liberated from thyroglobulin and one part of the hydrolyzed product (M^h, D^h) is enzymatically deiodinated with rate constant k^D, and the other part goes into the circulation with rate constant k^L. One part of the released iodide (I^-) is transported back into the epithelial cells and, together with other iodide ions, is reused. This is the intra-thryoidal iodine cycle. T_3^h and T_4^h are released into the blood with a rate constant k^R and transported to all parts of the body and used for hormone formation (T_3^c and T_4^c) with the rate constant k^E.

This flow chart is not a dogmatic assumption, but merely the starting point of an analogue computation which must start with certain values and assumptions.

Now let us assume that the quantity of doubly labeled molecules I*I* is very small and can be neglected. Then, considering the following rate processes

$$I^*I + T_y \xrightarrow{k_2} M^*, \quad I^*I + M \xrightarrow{k'_2} D^*,$$

$$I_2 + M^* \xrightarrow{k_3} D^*, \quad M^* + D \xrightarrow{k'_3} T_3^*, \quad M + D^* \xrightarrow{k'_4} T_3^*$$

we assume† the following relations:

$$k'_2 = k_3, \; k_2 = k'_2, \; k'_3 = k'_4 \tag{3.1}$$

† We cannot definitely rely upon such deductions, so that here we assume the relation (3.1) tentatively. We are considering another possibility too.

To obtain the fundamental equations of isotope kinetics we assigned numbers i to compartments as follows: Blood 0; thyroid follicle 1; I_2 in epithelial cells 2. Compartments between 0 and 1 or 1 and 2 are either neglected or the variables are grouped with 0, 1, or 2 for simplicity. The compartments corresponding to M^a, D^a, T_3^a, T_4^a are assigned 3, 4, 5, 6 and to M^i, D^i, T_3^i, $T_4^i - 3^i$, 4^i, 5^i and 6^i; to M^h, D^h, T_3^h, $T_4^h - 3^h$, 4^h, 5^h, 6^h; to 5^c, $6^c - T_5^c$, T_6^c respectively. I_i and e_i, are assumed equal to 0, ($I_i = 0$, $e_i = 0$, when $i \neq 0$) except that $i = 0 (I_0 \neq 0, e_0 \neq 0)$.

Let us write the system parameters of (1.5) as follows:

$$\left.\begin{array}{l} K_{01} = k_0;\ e_0 = k'_0 x_0;\ K_{10} = k'_1;\ K_{12} = k_1;\ K_{23} = k_2;\ K_{24} = k'_2; \\ K_{34} = k_3;\ K_{35} = k'_3;\ K_{45} = k'_4;\ K_{46} = k_4;\ K_{33^i} = K_{44^i} = K_{55^i} = \\ K_{66^i} = k^i;\ K_{3^i 3^h} = K_{4^i 4^h} = K_{5^i 5^h} = K_{6^i 6^h} = k^h;\ K_{3^h 1} = K_{4^h 1} = \\ k^D;\ K_{3^h 0} = K_{4^h 0} = k^L;\ K_{5^h 0} = K_{6^h 0} = k^R;\ K_{5^0 5^t} = K_{6^0 6^t} = k^E \end{array}\right\} (3.2)$$

Because iodine metabolism is in a steady state these parameters are constant in time.

The compartments are numbered according to the order of the flow chart shown in Fig. 1 and transitions of more than one step are as a rule neglected except for $2 \rightarrow 4$, $3 \rightarrow 5$ and $4 \rightarrow 6$. Transitions to compartments of smaller numbers are also omitted, except $1 \rightarrow 0$† ($K_{10} = k'_1$), and 3^h, $4^h \rightarrow 1$, ($K_{3^h 1} = K_{4^h 1} = k^h$). Then, as a trial, we can obtain the following simultaneous differential equations of 17 variables, (3.3) from the general phenomenological equation (1.5):

$$\left.\begin{array}{l} \dfrac{dx_0}{dt} = -(k_0 + k'_0)x_0 + k'_1 x_1 + I_0,\ \dfrac{dx_n^i}{dt} = -k^h x_n^i + k^i x_n \\ \qquad\qquad (n = 3, 4, 5, 6) \\ \dfrac{dx_1}{dt} = -(k_1 + k'_1)x_1 + k_0 x_0 + k^D(x_3^h + x_4^h) \\ \dfrac{dx_2}{dt} = -(k_2 + k'_2)x_2 + k_1 x_1,\ \dfrac{dx_n^h}{dt} = -(k^D + k^L)x_n^h + k^h x_n^i \\ \qquad\qquad (n = 3, 4) \\ \dfrac{dx_3}{dt} = -(k_3 + k'_3 + k^i)x_3 + k_2 x_2,\ \dfrac{dx_m^h}{dt} = -k^R x_m^h + k^h x_m^i \\ \dfrac{dx_4}{dt} = -(k_4 + k'_4 + k^i)x_4 + k'_2 x_2 + k_3 x_3 \\ \dfrac{dx_5}{dt} = -k^i x_5 + k'_3 x_3 + k'_4 x_4,\quad \dfrac{dx_m^c}{dt} = -k^E x_m^c + k^R x_m^h\ (m = 5, 6) \\ \dfrac{dx_6}{dt} = -k^i x_6 + k_4 x_4 \end{array}\right\} (3.3)$$

† k'_1 is the rate of back diffusion of iodine from the thyroid follicle to the circulation. The analogue computation considering such a diffusion or an abnormal leakage is to be attempted. The curve of Fig. 15 is only a preliminary one.

A hydraulic model showing these relations intuitively is shown in Fig. 2. Processes after x_5 or x_6 are omitted for simplicity.

If it is desired to change either the flow chart or the system parameters, (3.3) must be changed bearing in mind the general equations (1.5). For instance the re-use of the iodine of T_5^c and T_6^c are neglected in our treatment, although there is no concrete evidence for such an assumption. We therefore hope to test this larger iodine cycle by analogue computation in the near future. We have assumed, as a rule, that $k_1' = 0$, and only the computation shown in Fig. 15 took a value $k_1' \neq 0$ into consideration. This computation, however, is only preliminary and must be repeated.

The number of indeterminate parameters in (3.3) is too great to be determined by the trial and error method. Fortunately, as described in the next section, steady state considerations give useful data for their determination.

5. Steady State Considerations

Strictly speaking the concentration of an isotope must change from place to place even in the same compartment.† For simplicity, however, we assume that differences in concentrations in the same compartment can be neglected to a first approximation. Let us further assume that the frequency of turnover is sufficiently large so that equilibrium with respect to the ratio of the quantity of an isotope and that quantity of that element is always attained. Let us also assume that not only the entire system but also the distribution of the isotope itself is in a steady state. Then

$$\bar{x}_i/\bar{M}_i = \text{constant}$$

through all compartments, where $\bar{M}_i$ is the total quantity of all isotopes of the element and $\bar{x}_i$ is the steady state value of $\bar{x}_i$. Then

$$\bar{M}_0 : \bar{M}_1 : \bar{M}_2 \ldots = \bar{x}_0 : \bar{x}_1 : \bar{x}_2 \ldots \tag{4.1}$$

The mathematical foundation of this relation will not be described here (Fukuda & Sugita, 1959), but will be assumed intuitively.

The steady state value of x_i is the one obtained by putting $dx_1/dt = 0$ in (3.3). A mathematical consideration of this is given in another paper (Fukuda & Sugita, 1959).

From (3.3) significant statements can be made about the system parameters‡. For instance, in the steady state we have, with respect to the

† Although the compartment is uniform in respect to the stable compound, it may not be so in general with respect to the labeled compound which is added.

‡ If every steady state value of x_i with respect to the quantity of isotope were known, k_j, the system parameters, could be determined by using the relations of (3.3). Such a simple approach, however, is not available in our case. Only the method of trial and error may provide an approach.

quantity of isotope,

$$\dot{x}_2 = k_1\bar{x}_1 - (k_2 + k_2')\bar{x}_2 = 0$$
$$(k_2 + k_2')\bar{x}_2 = k_1\bar{x}_1 \tag{4.2}$$

Adding both sides of (3.3) pertaining to x_3, x_4, x_5, x_6 to each other, we have

$$\dot{x}_3 + \dot{x}_4 + \dot{x}_5 + \dot{x}_6 = (k_2 + k_2')x_2 - k^i(x_3 + x_4 + x_5 + x_6)$$

so that in the steady state

$$(k_2 + k_2')\bar{x}_2 = k^i(\sum_{n=3}^{6} \bar{x}_n) = k_1\bar{x}_1 \tag{4.3}$$

Again, from (3.3) pertaining to x_3^i, x_4^i, x_5^i, x_6^i we have in the steady state

$$\dot{x}_n^i = k^i\bar{x}_n - k^h\bar{x}_n^i = 0, \ (n = 3, 4, 5, 6) \tag{4.4}$$

Then

$$\bar{x}_n^i = \frac{k^i}{k^h}\bar{x}_n \text{ and } \sum_{n=3}^{6} \bar{x}_n^i = \frac{k^i}{k^h}\sum_{n=3}^{6} \bar{x}_n \tag{4.5}$$

From (4.3) and (4.5) we have one of the *important relations*

$$\sum_{n=3}^{6} \bar{x}_n^i = \frac{k_1}{k^h}\bar{x}_1 \tag{4.6}$$

From (4.1) and experimental data on the net metabolism of all isotopic forms of iodine we can assume as a reasonable value (Suzuki, 1958) of the ratio

$$\bar{x}_1 / \sum_{n=3}^{6} \bar{x}_n^i = \bar{n}_1(\mathrm{I}_2) / \sum_{n=3}^{6} \bar{n}_n^i(\mathrm{I}_2) = 1/100$$

Therefore, from (4.6) we shall assume, as an approximation

$$k^h = 0{\cdot}01k_1 \tag{4.7}$$

Again, from (3.3) we have in the steady state

$$\dot{x}_1 = -(k_1 + k_1')\bar{x}_1 + k_0\bar{x}_0 + k^D(\bar{x}_3^h + \bar{x}_4^h) = 0 \tag{4.8}$$
$$(k_1 + k_1')\bar{x}_1 = k_0\bar{x}_0 + k^D(\bar{x}_3^h + \bar{x}_4^h)$$

Also, in the steady state

$$\dot{x}_n^h = k^h\bar{x}_n^i - (k^D + k^L)\bar{x}_n^h = 0, \ (n = 3, 4)$$

so that

$$(k^D + k^L)(\bar{x}_3^h + \bar{x}_4^h) = k^h(\bar{x}_3^i + \bar{x}_4^i) \tag{4.9}$$

From (4.6) and (4.9) we have in cancelling k^h

$$k^D(\bar{x}_3^h + \bar{x}_4^h) = \alpha\frac{\bar{x}_3^i + \bar{x}_4^i}{\sum_n x_n^i} \ k_1\bar{x}_1, \tag{4.10}$$

where

$$\alpha = k^D/(k^D + k^L). \tag{4.11}$$

Therefore, from (4.8) and (4.10) we have the *second important equation*

$$(k_1 + k_1')F^{-1}\bar{x}_1 = k_0\bar{x}_0 \tag{4.12}$$

where

$$F^{-1} = 1 - \alpha \frac{\bar{x}_3^i + \bar{x}_4^i}{\sum_{n=3}^{6} \bar{x}_n^i} \frac{k_1}{k_1 + k_1'} \tag{4.13}$$

Observational data on the ratio of $\bar{n}_0(I_2)$ and $\bar{n}_1(I_2)$ indicate 1/10 as an approximation, so that we shall assume the value

$$\bar{x}_1 = 10\bar{x}_0 \text{ and } \frac{k_0}{k_1 + k_1'} F = 10 \tag{4.14}$$

The approximate ratios of $\bar{n}_3^i(x)$, $\bar{n}_4^i(I_2)$, $\bar{n}_5^i(I_2)$ and $\bar{n}_6^i(I_2)$ is known to be (Suzuki, 1958)

$$\bar{x}_3^i : \bar{x}_4^i : \bar{x}_5^i : \bar{x}_6^i = \bar{n}_3^i(I_2) : \bar{n}_4^i(I_2) : \bar{n}_5^i(I_2) : \bar{n}_6^i(I_2) = 40 : 47 : 2 : 12$$

so that

$$\frac{\bar{x}_3^i + \bar{x}_4{}^i}{\sum_n \bar{x}_n^i} = \frac{87}{101} \tag{4.15}$$

If, on the one hand, we put $k_1' = 0$ and $k^L = 0$ (Taurog, Wheat & Chaikoff, 1955) then

$$F^{-1} = 1 - \frac{87}{101} = \frac{13}{101},$$

and

$$\frac{k_0}{k_1} F = \frac{k_0}{k_1} \frac{101}{13} = 10$$

Therefore, the approximation of k_0 is

$$k_0 = k_1 \tag{4.16}$$

If, on the other hand, we put $k_1' = 0$, $k^L \neq 0$ and assume the value, as a trial

$$\alpha = \frac{k^D}{k^D + k^L} = \frac{4}{5}$$

then

$$F^{-1} = 1 - \frac{348}{505} = \frac{1570}{505} \tag{4.17}$$

and

$$k_0 = 10F^{-1}k_1 = \frac{1570}{505} k_1 \tag{4.18}$$

Then the approximation of k_0 can be written as

$$k_0 = 3k_1 \tag{4.16'}$$

From (3.3) and (4.5) we have the relations

$$\frac{\bar{x}_4}{\bar{x}_6} = \frac{\bar{x}_4^i}{\bar{x}_6^i} = \frac{k_4}{k^i} = \frac{12}{47}, \quad \therefore k^i = 4k_4 \tag{4.19}$$

$$\frac{\bar{x}_4}{\bar{x}_3} = \frac{\bar{x}_4^i}{\bar{x}_3^i} = \frac{k_2' + k_2k_3'(k_3 + k_3' + k^i)}{k_2}\left(\frac{k_3 + k_3' + k^i}{k_4 + k_4' + k^i}\right) = \frac{47}{40} \tag{4.20}$$

$$\frac{\bar{x}_5}{\bar{x}_6} = \frac{\bar{x}_5^i}{\bar{x}_6^i} = \frac{k_3'(\bar{x}_3/\bar{x}_4) + k_4'}{k_4} = \frac{k_3' + k_4'(47/40)}{k_4(47/40)} = 1/6 \tag{4.21}$$

Then, from (3.1) and (4.21) we have as approximations of k_3' and k_4'

$$k_3' = k_4' = \frac{1}{10}k_4 \tag{4.22}$$

and from (4.20)

$$\left(1 + \frac{k_3}{k_3 + k_3' + k^i}\right)\left(\frac{k_3 + k_3' + k^i}{k_4 + k_4' + k^i}\right) = \frac{47}{40}$$

and from (4.19), (3.1) and (4.22)

$$\frac{2k_3 + k_3' + k^i}{k_4 + k_4' + k^i} = \frac{2}{5{\cdot}1}\frac{k_3}{k_4} + \frac{41}{51} = \frac{47}{40}$$

$$\therefore \frac{k_3}{k_4} = \frac{5{\cdot}1}{2}\left(\frac{47}{40} - \frac{41}{51}\right) \doteqdot 1$$

Therefore, the approximation of k_4 is

$$k_4 = k_3 \tag{4.23}$$

Taking into consideration the relations (4.7), (4.16), or (4.16′), (4.19), (3.1), (4.22), (4.23) and assuming $k_1' = 0$, we can put

$$k_2 = k_2' = k_3 = k_4 = 10k_1, \; k_3' = k_4' = k_1, \; k^i = 40k_1, \; k^h = 0{\cdot}01k_1$$

and

$$k_0 = k_1 \text{ for } k^L = 0, \; \alpha = 1$$

$$k_0 = 3k_1, \text{ for } k^L \neq 0, \; \alpha = 4/5 \tag{4.24}$$

In the latter case we put, temporarily,

$$k^D = 40 \text{ and } k^L = 10 \tag{4.25}$$

as plausible values. From these approximate values we can compute the following reasonable values

$$\bar{x}_3^i = 39{\cdot}25\bar{x}_1, \; \bar{x}_4^i = 46{\cdot}67\bar{x}_1,$$

$$\bar{x}_5^i = 2{\cdot}15\bar{x}_1, \; \bar{x}_6^i = 11{\cdot}67\bar{x}_1,$$

$$\text{and } \bar{x}_1 = 9{\cdot}65\bar{x}_0$$

We are currently uncertain of the experimentally determined value of k^R so that temporarily we put

$$k^R = 10k_1 \tag{4.26}$$

For the determination of k^E we shall use the relation

$$\bar{x}_5^e + \bar{x}_6^e = \frac{k^R}{k^E}(\bar{x}_5^h + \bar{x}_6^h) = \frac{k^h}{k^E}(\bar{x}_5^i + \bar{x}_6^i) =$$

$$\frac{k^h}{k^E}\left(\sum_{n=3}^{6} \bar{x}_n^i - \bar{x}_3^i - \bar{x}_4^i\right)$$

From (4.6), (4.12) and (4.13) we have

$$\bar{x}_5^e + \bar{x}_6^e = \frac{k_1}{k^E}\left(1 - \frac{1}{\alpha}\frac{k_1 + k_1'}{k_1}\frac{F-1}{F}\right)\frac{Fk_0}{k_1 + k_1'}\bar{x}_0 =$$

$$\frac{k_0}{k^E}\left(F - \frac{1}{\alpha}\frac{k_1 + k_1'}{k_1}(F-1)\right)\frac{k_1}{k_1 + k_1'}\bar{x}_0 \tag{4.27}$$

This is the *third important relation.*

If we assume $k_0' = 0$, $\alpha = 4/5$ and

$$k^E = 0{\cdot}3k_1 \tag{4.28}$$

then, from (4.27) we can calculate the reasonable value

$$\bar{x}_5^e + \bar{x}_6^e = 4{\cdot}77\bar{x}_0$$

The determination of these values for the system parameters is merely a trial, so it is desirable to test other possibilities. Due to lack of time we have not done this, but expect to do so in the future using an analogue computer.

That we can obtain reasonable values of system parameters and get relations like (4.6), (4.12) and (4.27) from steady state considerations is significant and may be of use in medical practice.

6. Use of the Analogue Computer

We have used the analogue computer at Meiji University with the kind cooperation of Professor Jinbo (Jinbo, Ogawa & Imura, 1960). The block diagram is shown in Fig. 3, when $k_1' = 0$. When $k_1' \neq 0$, the block diagram is different from Fig. 3.

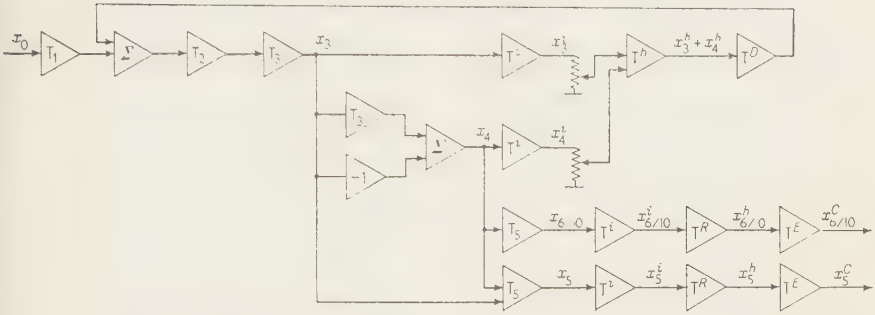


FIG. 3. Block diagram for the analogue computation of iodine metabolism.

$$T_1 = \frac{k_0/k_1}{1 + p/k_1}, \qquad T^D = \frac{k^D/k_1}{1 + p/k_1}, \qquad k'_1 = 0,$$

$$T_2 = \frac{k_1/(k_2 + k'_2)}{1 + p/(k_2 + k'_2)}, \qquad T^h = \frac{k^h/(k^D + k^L)}{1 + p/(k^D + k^L)},$$

$$T_3 = \frac{k_2/(k_3 + k'_3 + k^i)}{1 + p/(k_3 + k'_3 + k^i)}, \qquad T^R = \frac{k^h/k^R}{1 + p/k^R},$$

$$T_5 = \frac{k'_4/k^i}{1 + p/k^i}, \qquad T^i = \frac{k^i/k^h}{1 + p/k^h}, \qquad T^E = \frac{k^R/k^E}{1 + p/k^E}.$$

When $k'_1 = 0$, instead of applying† the input pulse $I_0(t)$, we have used another analogue circuit, the solution of which is

$$x_0 = Qe^{-(k_0 + k'_0)t} \tag{5.1}$$

and the output x_0 is applied as the input. Q of (5.1) corresponds to the injected quantity of ^{131}I.

The curves that were obtained are reproduced in Figs. 4 to 18. Where necessary, a note (like CV, 3/8) indicates that the true computed value (CV) referred to x_1 is 3/8 times the apparent height of the curve. In the case of x_1 CV is referred to Q. The pen recorder magnification is written as, say, PRM, 4.

In Fig. 4 the value of x_1 shows a sharp peak, while the peak of the pulse of (5.1) is too steep to have an appreciable breadth on this time scale. (Time marks correspond to 1-second intervals.) Upper and lower curves on the same graph are recorded simultaneously. The lower step curve is the input to obtain the output x_0 of (5.1), the height of which is Q.

The sharp peak in Fig. 5 of x_3 or x_4 is similar to x_1. However, its height is only about 1/100 that of x_1, which explains why compounds M and D cannot be found in the epithelial cells although this is where the incorporation of I_2 is assumed to take place. This is demonstrated visually by the hydraulic analogue in Fig. 2. The level in the tank having a large k_1 is

† It is better to apply step input to the operational amplifier next to the integrator of $\dot{x}_0$ designing the block diagram different from that of Fig. 3.

very much lower than in the tank having a small k_1. For the same reason the x of a compartment having a sharp peak (or short life) must be very

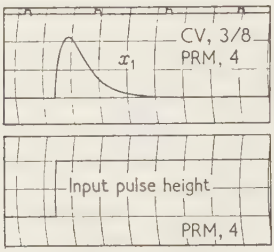


FIG. 4.

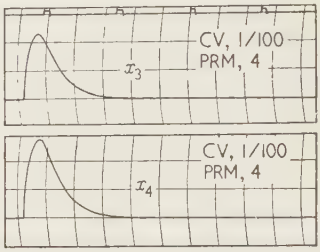


FIG. 5.

FIG. 4. The sharp peak representing the concentration of ^{131}I transported into the thyroid follicle. $k_0 = 3$

FIG. 5. The peaks corresponding to M and D in the epithelial cells. $k_0 = 3$

small, so that it may not be detectable experimentally. We believe this explains why it has been believed that I_2 is incorporated at some place other than the epithelial cells. However, our computations appear to justify our assumption of incorporation in these cells.

Fig. 6 shows the relation between x_1 and x_3^i. The curve of x_3^i does not show a sharp peak like that of x_1. In this case $k_0 = 1$ instead of $k_0 = 3$.

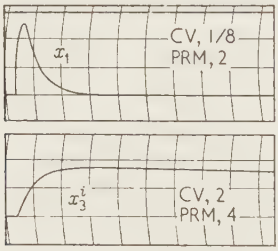


FIG. 6.

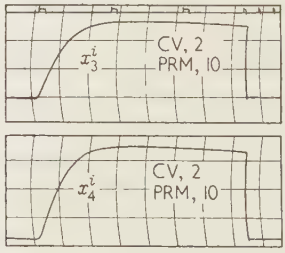


FIG. 7.

FIG. 6. Comparison of the peaks of I^- and M in the thyroid follicle. $k_0 = 1$

FIG. 7. Upper curve, M in the thyroid follicle; lower curve, D in the thyroid follicle. $k_0 = 3$; $k^h = 1/100$; $k'_1 = 0$; $k'_0 = 3$

The upper and lower curves of Fig. 7 are those of x_3^i and x_4^i, respectively, and the upper curves of Fig. 8 and Fig. 9 those of x_5^i and x_6^i. From these curves we observe that the height of x_3^i is almost the same as $8x_1$,

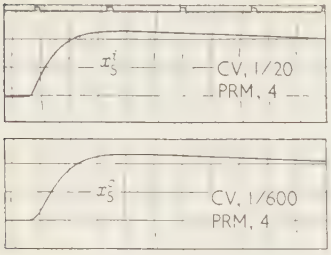


FIG. 8.

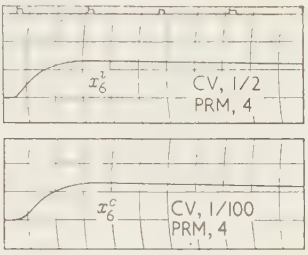


FIG. 9.

FIG. 8. Upper curve, T_3 in the thyroid follicle; lower curve, ^{131}I-labelled triiodothyronine.
$k_0 = 3$; $k^h = 1/100$; $k'_1 = 0$; $k'_0 = 3$

FIG. 9. Upper curve, T_4 in the thyroid follicle; lower curve, ^{131}I-labelled thyroxnine.
$k_0 = 3$; $k^h = 1/100$; $k'_1 = 0$; $k'_0 = 3$

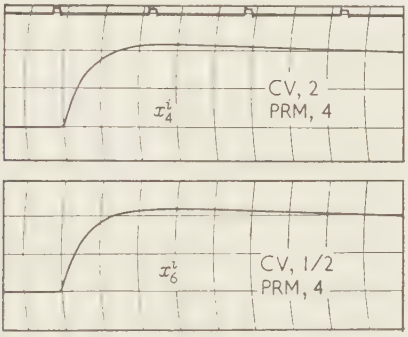
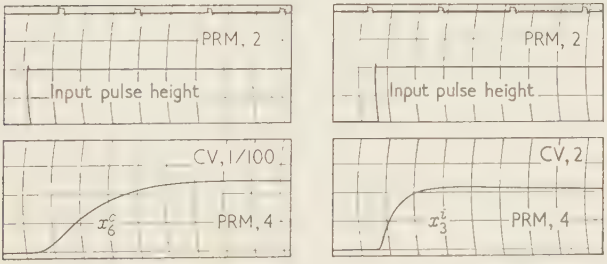


FIG. 10. Behavior of D and T_4 in the thyroid follicle.
$k_0 = 1$



(a) (b)

FIG. 11 (a). Time lag of the synthesis of thyroxine.
$k_0 = 0{\cdot}1$

FIG. 11 (b). Small time lag of the production of D^i.
$k_0 = 0{\cdot}1$

and that $x_4^i > x_3^i$, as expected. x_6^i is only about 1/5 of x_4^i (Fig. 10). When compared with the results of observation these results are quite reasonable. The lower curve of Fig. 8 shows the curve of x_5^c. x_6^c is about 1/100 of x_1 (Fig. 9), again not very different from the results of observation.

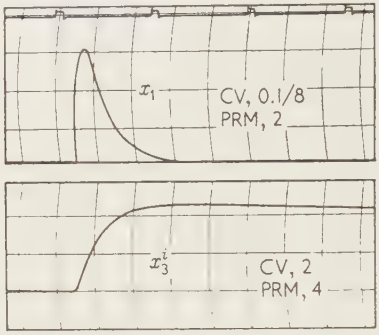


FIG. 12. Comparison of the peaks of I^- and M^i. Fig. 12 corresponds to Fig. 6. $k_0 = 0{\cdot}1$; $k^h = 1/100$

In Fig. 11a the curves are recorded simultaneously with the input step, and demonstrate a time lag in the initial ascension of the curve of x_6^c. In Fig. 11b the curve of x_3^i does not show a noticeable delay. $k_0 = 0{\cdot}1$ in Fig. 11 and Fig. 12, and the curve of Fig. 12 corresponds to Fig. 6. A

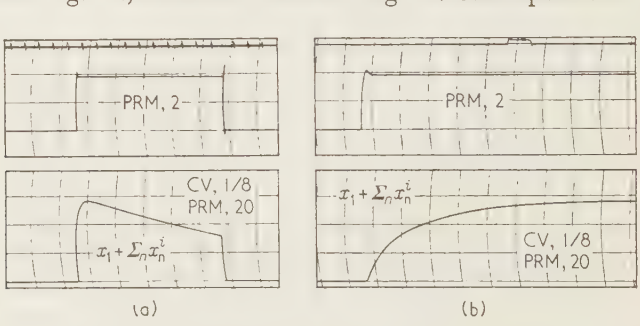


FIG. 13 (a) and (b). Uptake curve of ^{131}I (normal). $k_0 = 1$; $k^h = 1/100$; $k'_1 = 0$; $k'_0 = 3$
The time interval of (a) is 1/5 the normal one and that of (b) is 5 times the normal.

value of $k_0 = 0{\cdot}1$ may correspond to the case of a small animal, but $k'_0 = 3$ surely does not. This may indicate, perhaps, a pathological condition in man.

Figs. 13, 14 and 15 show uptake into the thyroid gland, obtained by summing the outputs of x_1, x_3^i, x_4^i, x_5^i and x_6^i. In a and b of each figure the time interval is respectively 1/5 and 5 times that normally used.

Except for Fig. 6, we have taken $k_0 = 3$ in most of the figures 4 to 18. $k_0 = 1$ in Figs. 10, 13, 14, 15. The value of $k_0 = 3$ is too high for normal

Japanese, so that we took $k_0 = 1$. In this case the maximum uptake can be estimated as about 20%, an acceptable normal value. As a reference, we present the case of $k_0 = 0{\cdot}1$ in Figs. 11, 12.

In Fig. 13, $k^h = 0{\cdot}01$, $k'_1 = 0$, in Fig. 14, $k^h = 0{\cdot}1$, $k'_1 = 0$, in Fig. 15 $k^h = 0{\cdot}01$, $k'_1 = 3$, $k'_0 = 5$. PRM is 2 for the step input of all these curves.

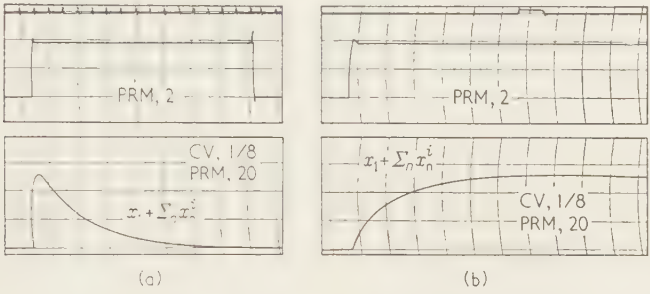


FIG. 14 (a) and (b). Uptake curve of ^{131}I (abnormal).
$k_0 = 1$; $k^h = 1/10$; $k'_1 = 0$; $k'_0 = 3$
The time interval for (a) is 1/5 the normal one and that of (b) is 5 times the normal.

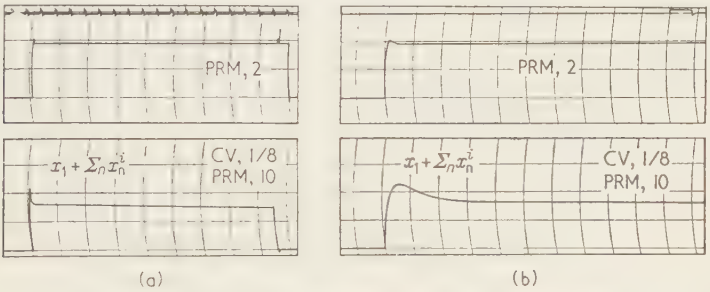


FIG. 15 (a) and (b). Uptake curve of ^{131}I (abnormal).
$k_0 = 1$; $k^h = 1/100$; $k'_1 = 3$; $k'_0 = 5$
The time interval for (a) is 1/5 the normal and that of (b) is 5 times the normal.

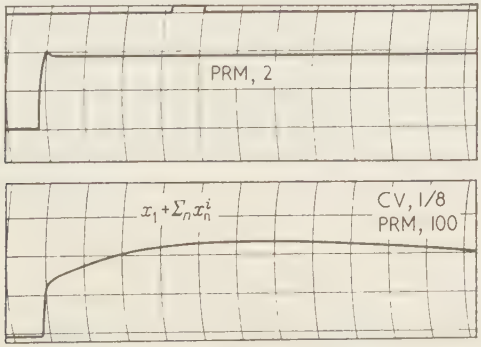


FIG. 16. Uptake curve of ^{131}I (abnormal).
$k_0 = 3$; $k'_0 = 3$; $k^h = 1/10$; $k'_1 = 0$
The time interval is 5 times the normal.

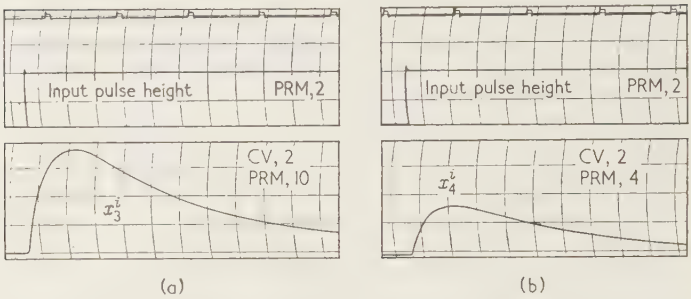


FIG. 17 (a). Behavior of M^i in the thyroid follicle (abnormal).
$k_0 = 1$; $k^h = 0{\cdot}1$; $k'_1 = 0$

FIG. 17 (b). ^{131}I-labelled D^i (abnormal).
$k_0 = 1$; $k^h = 0{\cdot}1$; $k'_1 = 0$

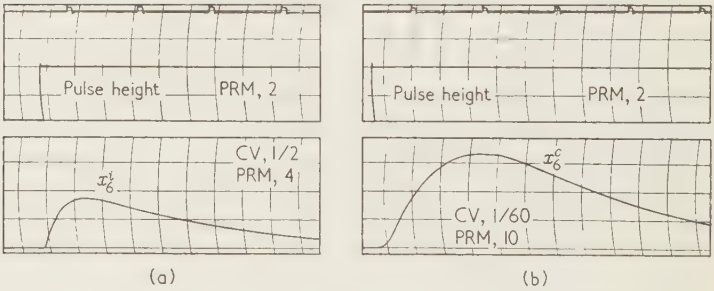


FIG. 18 (a). Behavior of T_4 in the thyroid follicle (abnormal).
$k_0 = 1$; $k^h = 0{\cdot}1$; $k'_1 = 0$

FIG. 18 (b). ^{131}I-labelled thyroxine (abnormal).
$k_0 = 1$; $k^h = 0{\cdot}1$; $k'_1 = 0$

If we assume a large value of hydrolysis, $k^h = 0{\cdot}1$, the computed curve shows a steep descent (Fig. 14), $k'_1 = 0$ in all curves except that of Fig. 15. We used the value $k'_0 = 5$ through an error in adjusting the computer. An initial peak and a subsequent plateau can be noted. Fig. 16, in which $k_0 = 3$, corresponds to Fig. 14b. Curves of x_3^i and x_4^i are shown in Fig. 17 and those of x_6^i and x_6^c are shown in Fig. 18. Those curves show a steep descent compared with those of Figs. 7, 8 and 9.

Curves of Fig. 13, 14, 15 were obtained by summing the computed values of the outputs corresponding to these curves. (x_5^i was not recorded at that time.)

6. Conclusions

A mathematical formulation of *in vivo* processes is so complex that numerical analysis is hardly possible by ordinary methods. Analogue computations provide a powerful tool in evading such mathematical

difficulties. Even so, the accuracy of an analogue computer may not be sufficient and the use of digital computers or digital differential analysers may be required. We intend to use the latter in the future. However, the application of the analogue computer has an important influence on the manner of thinking in approaching such problems.

Before the development of methods for computing by machine, it was possible to approach such problems only when both the starting equations and parameters or numerical data had been obtained experimentally. Now, however, it is possible to compute on a trial basis, so that one can search or, so to speak, synthesize equations which suitably describe the behavior of a living system. If reasonable results are obtained this may be taken as indicating that the equations correctly describe the system, also they are not the only one. Some may as yet feel that our approach lacks a sufficient experimental basis. However, the computations themselves sometimes provide such a basis. The biology of the future must be extended using analogue as well as digital computers. We plan to repeat and extend our computations, since this approach may be very useful for analysis of metabolism as well as research in the field of endocrinology.

The senior author presented our results at a meeting of the Endocrinological Society held at Osaka (April, 1960) where the possibility of verifying our results experimentally was discussed.

The authors are grateful to Professor Jinbo for his special interest, and also for the cooperation of Messrs. Y. Ogawa and Y. Hirabara, all of the Laboratory of Electrotechnology, Meiji University.

REFERENCES

Chance, B. (1955). *J. biol. Chem.* **217,** 429.
Chance, B. & Hess, B. (1959). *Science,* **129,** 700.
Fukuda, N. & Sugita, M. (1959). *Bull. Kobayasi Inst.* **9,** No. 3–4, 1. In Japanese.
Fukuda, N. & Sugita, M. (1960). *Bull. Kobayasi Inst.* **10,** No. 1–2, 4. In Japanese.
Hess, B. & Chance, B. (1959). *Naturwissenschaften,* **46,** 284.
Jinbo, S., Ogawa, Y. & Imura, S. (1960). *Rep. Fac. Engn. Meiji Univ.* **1.** In Japanese.
Sugita, M. (1954). *Bull. Kobayasi Inst.* **4,** No. 3–4, 199. In Japanese.
Sugita, M. (1960). *Buturi,* **15,** No. 10, 528. Review in Japanese.
Suzuki, M. (1958). *Kagakuno Ryoiki,* **12,** 715. Review in Japanese.
Taurog, A., Wheat, J. D. & Chaikoff, I. L. (1955). *Trans. Amer. Goiter Assoc.* 228.

J. Theoret. Biol. (1961) 1, 460–487.

A Simulation Study of a Diffuse Conducting System based on Coelenterate Nerve Nets

Robert K. Josephson

Zoology Department, University of California, Los Angeles, California, U.S.A.

Richard F. Reiss and R. M. Worthy

Librascope Division, General Precision Inc., Glendale, California, U.S.A.

(*Received 22 May* 1961)

A model of a coelenterate nerve net has been constructed for digital computer simulation experiments. The properties of the model were chosen so as to be as consistent as feasible with those known for coelenterate nets. Experiments with the model offer a means of testing the adequacy of current physiological concepts of excitation-spread in nerve nets and can show what sorts of behavior should be expected of such nets. Among the findings of this study which are pertinent to an understanding of spread in living nerve nets which are not through-conducting are the following:

1. Unless repetitive firing is a factor it is to be expected that nets that conduct for considerable distances relative to the lengths of their conducting elements following a stimulus will also give quite variable responses. Factors that lead to a greater relative distance of spread, a larger ratio of transmissive to non-transmissive junctions between conducting elements or a larger number of junctions per conducting element, also lead to greater variability in the distance of spread.

2. Increasing the stimulus strength (increasing the number of elements stimulated) should increase the distance of spread. But as more elements are stimulated, occlusion of possible pathways becomes an important factor and tends to limit the effectiveness of increasing stimulus strength.

3. Spread following repetitive stimulation can be expected to occur in equal increments of distance following each stimulus or in decreasing increments of distance; spread in increasing increments requires postulates not included in the model.

4. On the basis of excitation-spread following a single stimulus or repetitive stimulation of the same area, one can not expect to distinguish between a single non-polarized junction at each neuron crossing and two reciprocally polarized junctions, or between one-way facilitation and two-way facilitation. Sequential stimulation of two points whose response areas slightly overlap is an experiment which might distinguish between these possibilities.

Introduction

A diffuse, randomly organized, monoplanar network of nerve cells is perhaps the simplest assemblage of interacting neurons which can be imagined. Yet such nerve nets form almost the entire nervous system of all coelenterates, and there is histological evidence that they form parts of the peripheral nervous systems of many other animal groups.

There is evidence that the neurons and the elementary nervous events of such nets are in many ways similar to those of the higher, more studied animals. Horridge (1954a) demonstrated that conduction is associated with the usual all-or-none action potential (see also Yamashita, 1957; Passano, 1958; Passano and McCullough, 1960; Josephson, 1961b). Some of the earlier workers (Bethe, 1903; Parker, 1919) believed that the neurons of coelenterate nerve nets are fused to form a syncytium. More recent work by Pantin (1952), however, has convincingly shown that there is contiguity rather than continuity between neurons, at least in the mesenteric net of the sea anemone *Metridium*, and the same seems to be true of the nets of scyphozoan medusae (Bozler, 1927) and in certain nets of siphonophores (Mackie, 1960). In the latter group, both syncytial and non-syncytial nets co-exist in the same tissue. In non-syncytial nets, interaction between neurons probably occurs where these neurons meet in crossing, and these junctions may be properly called synapses. It is with such non-syncytial nets that this paper deals.

Coelenterate nerve nets, although apparently randomly organized, are nevertheless capable of some integrative activity. In corals and other colonial coelenterates, the individual members of the colony are joined by what appears to be a nerve net. Stimulation of a portion of a colony evokes retraction of all members of the colony in some species, while in others only a local patch of polyps responds (Parker, 1920; Horridge, 1957; Josephson, 1961a). In the latter species, the size of the responding area usually increases with increasing number of stimuli and with increasing stimulus strength, especially following mechanical stimulation but often also following brief electric shocks. Thus the number of responding polyps, that is, the distance to which excitation spreads, is a function of the stimulus. Similar responses have been found in parts of individual coelenterates, for example, the oral disk of sea anemones (Pantin, 1935a). These properties are dealt with in different ways in different models (Pantin, 1935a; Horridge, 1957) and are in part not yet adequately accounted for.

Before experiments on systems containing nerve nets can be fully evaluated, one must know what properties to expect from such nets. Activity in a diffuse system could involve interaction between a large

number of elements, and the results to be expected from such interaction are by no means obvious. Repetitive firing, even following brief electric shocks, has been shown to account for some of the observed properties of excitation-spread in hydroid colonies (Josephson, 1961b). This does not solve the problem, however, because repetitive firing in a nerve net can only operate within the framework of possibilities offered by the net.

Horridge (1957) proposed two models of nerve nets, one mechanical, the other mathematical. Using the mechanical model, he studied the spread of excitation following "stimulation" at one point in a network made of elements joined by some transmissive and some non-transmissive junctions and in which non-transmissive junctions could be made transmissive by previous activity. The results of experiments on this model were too varied to accept it as an explanation of coelenterate behavior; that is, spread was very different in repeated tests. The mathematical model was an attempt to reduce the variability by allowing many units to be initially activated simultaneously. Using this model, Horridge was able to explain the greater spread which followed more intense or repeated stimuli as due to more elements being initially activated and, thus, there being a greater probability of the excitation finding long transmitting pathways.

Horridge's second model, though it does explain many of the observed characteristics of spread in coelenterates, has three major weaknesses. First, great emphasis is placed upon an increasing number of elements being initially activated by each stimulus of a series—a proposition for which there is little supporting evidence; and interneural facilitation—a factor which one expects, *a priori*, to be quite important—is given little significance. Second, while Horridge uses the term *density* in his equations, they actually apply to the *number* of units active. Knowing the number of units active after a number of junctions have been crossed is not equivalent to knowing the density of active units (a quantity which may fall as the area increases with spread), and is not equivalent to knowing the distance to which excitation will spread, since the percentage of active elements in sub-areas may fall in an unspecified manner. Finally, Horridge's equations, which are maximum likelihood estimates for a series of trials with each trial being made up of a number of individual events, demand independence between the individual events forming a single trial. The probability that excitation in one element will influence another cannot be a function of the number or density of active elements. This is perhaps not the case in a nerve net where occlusion of possible pathways is probably significant, especially with higher densities of excited units.

Simulation seems a logical approach to the study of the behavior to be

expected of nerve nets of defined structure and properties when there is no clear way of evaluating analytically the effects of such factors as interneural facilitation, the increasing area encompassing the active elements as excitation spreads from the origin, or the interaction by occlusion of activated elements. It was for this reason that the present study was instigated. It is also hoped that this report will show, by example, the applicability of digital computer simulation to such biological problems.

2. The Simulation Model

The properties of the simulation model were carefully chosen to be as consistent as feasible with known histological and physiological properties of coelenterate nerve nets. Two important features incorporated in the model were first proposed by Horridge (1957): (1) the initial mixed population of transmissive and non-transmissive junctions linking conducting elements; and (2) the simultaneous "stimulation" of a variable number of elements to start excitation in the net. The considerations leading to these ideas and their bearing on the model are discussed below.

The model was first programmed for an LGP-30, a relatively small digital computer. After initial testing and some preliminary experiments, it was recoded for a larger computer, the IBM 709, and most of the data to be presented were obtained from experiments on this machine.

For convenience, description of the simulation model may be divided into three parts: (1) the net topology—the arrangement of the intersecting elements which form the nerve net; (2) the functioning of the crossings—those places where one element contacts another and where interaction between the two elements can occur; and (3) the method by which spread of excitation through the net is simulated.

A. THE NET TOPOLOGY

The coelenterate nerve net often resembles an array of "pick-up-sticks" or "jack-straws" of varied lengths, strewn about a surface with more or less uniform density (see Fig. 1 for some examples of coelenterate nerve nets). This analogy emphasizes that the conducting elements are straight, and only touch; they do not fuse with one another. A topology like that of randomly strewn sticks, however, would be difficult to duplicate, even on a high-speed computer. This distribution is approximated in the model by a network similarly made of linear conducting units, but of units that always intersect at right angles.

The topology of the simulation net may be viewed as a square grid of horizontal and vertical lines in which certain segments have been removed, leaving a number of horizontal and vertical segments of varied lengths as shown in Fig. 2. The model, therefore, lacks side-by-side or end-to-end

connections between parallel elements. This is not in significant disagreement with known histology.

In the production of the net, the frequency distribution of lengths of conducting elements can be determined by the experimenter, but the sequence in which these element lengths appear in any horizontal or

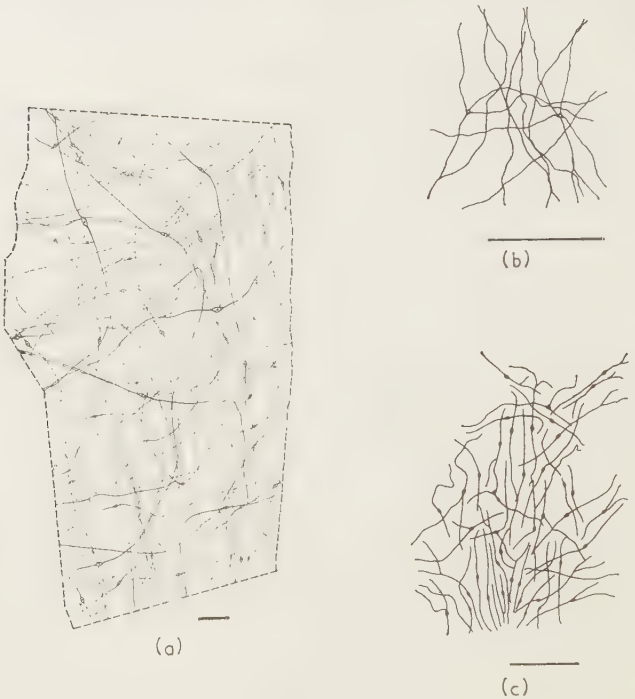


FIG. 1. Some examples of coelenterate nerve nets. The bar in each case represents 0·2 mm.

(*a*) The net in the mesentery of *Metridium* (from a preparation and unpublished drawing of Dr. E. Batham). For clarity, some of the neurons are shown as dotted lines.

(*b*) The non-syncytial net in the external ectoderm of *Velella* (redrawn, omitting the syncytial net, from Mackie, 1960).

(*c*) A diagram of part of the subumbrella net of *Cyanea* (redrawn from Horridge, 1954b).

vertical line is randomly determined. For example, if it were decided that three-fourths of the conducting elements of the net should be four units long and the other fourth five units long (a unit being the distance between adjacent parallel lines in the original grid), after removing one segment in the grid the computer would randomly decide whether to skip four segments or five segments before removing another, using an algorithm such that the probability of the first outcome would be three times as great as that of the second.

The net topology is "stored" in the computer by means of the following device: each point in the net where two conducting elements cross is uniquely assigned a memory cell in the computer. A portion of each of these memory cells is devoted to a description of the connections between its corresponding crossing and the four neighbors—the crossings above,

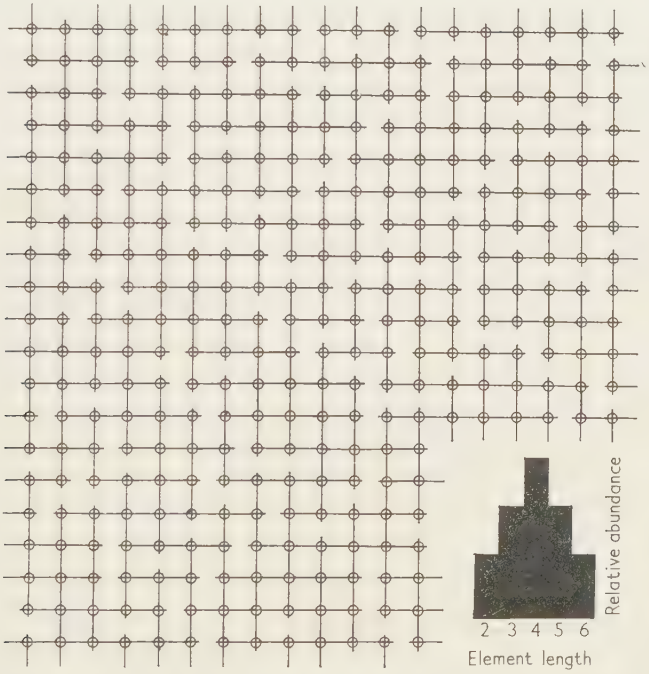


FIG. 2. A portion of a typical simulation net, showing the rectangular array of horizontal and vertical conducting elements of varying lengths. The circled intersections represent the crossings where interaction between the elements occurs. The histogram shows the frequency distribution of lengths of the conducting elements in the whole net from which this portion was taken.

below, to the right, and to the left. For example, if crossing x is immediately above crossing y in the network, and if a vertical element runs through both crossings (i.e. they are "connected") there will be a symbol in the cell representing x which indicates that x is connected to y, and similarly, a symbol in y indicating a connection with x.

The nets created for these experiments were square; 32 units on a side in the experiments done on the LGP-30, 100 units on a side in those done on the IBM 709. The frequency distribution of lengths of conducting elements used was always normal, and the average element was four units long in the majority of the nets used in experiments. Such nets in the

LGP-30 experiments were composed of about 385 elements and those in the IBM 709 experiments of about 4,000 elements.

B. THE FUNCTIONING OF THE CROSSINGS

Activity in one conducting element can induce activity in another only where these two elements intersect. Obviously the properties of these crossings will greatly affect the spread of activity throughout the net.

The properties of junctions in a living nerve net can be inferred from the results of experiments on such nets. In the discussion below, special attention will be given to nerve nets which are not through-conducting. The best data are from investigations on nerve nets linking individuals in colonial animals (Horridge, 1957; Josephson, 1961a).

In evaluating such experiments, the assumption is commonly made that excitation is conducted in the coelenterate nerve net by all-or-none impulses. Evidence for such an assumption is given by the work of Pantin (1935a) and, more especially, the direct observations of Horridge (1954a) who first found all-or-none action potentials in a scyphozoan medusae. Further, the great distance of spread in many nerve nets precludes the possibility of decremental conduction. In the following discussion, it will be assumed that activity in the conducting units is all-or-none and that decremental activity plays no role at this level.

Conduction in a coelenterate nerve net is classically non-polarized; i.e. it proceeds equally well in any direction. Even individual neurons in the net can conduct normally in either direction (Horridge, 1954a).

There would seem to be two possible ways to explain non-polarized transmission between neurons. Each neuron might be linked with other neurons by connections which are themselves non-polarized and transmit excitation either to or away from the neuron. Alternatively, the neurons might be linked only by polarized connections, but with each neuron having some connections that transmit excitation to the neuron and some which transmit excitation away from the neuron. There is little physiological evidence to allow choice between these possibilities. The occasional occurrence of polarized conduction in coelenterates (Rand, 1909; Pantin, 1935b; Horridge, 1956) would indicate polarized connections can occur. On the other hand, the beautiful histological preparations of Batham (Pantin, 1952), which reveal, at least at the level of light microscopy, no asymmetry on either side of the intersection of neurons in the mesenteric nerve net of the sea-anemone *Metridium*, would tend to indicate non-polarized connections. It seems likely that both situations can occur in different nerve nets or in different parts of the same net.

In the model, excitation can be transmitted in either direction at the intersections between conducting elements. These crossings, however,

are divided into two parts for flexibility of simulation experiments, one transmitting excitation from the horizontal conducting element to the vertical element, the other transmitting in the opposite direction. That portion which transmits from horizontal to vertical is considered part of the horizontal element; that which transmits from vertical to horizontal, part of the vertical element. For lack of a better term, these two parts will be called "junctions" in discussing the model. Each crossing, therefore, contains two junctions, each polarized for transmission in one direction.

In all cases where the distance of spread of excitation is greater than the length of any single conducting unit, and this must often be the case in the polyp withdrawal response in coral colonies (Horridge, 1957), several conducting units must have been serially activated and, barring the possibility of repetitive firing evoked by the stimulus, must be linked by transmissive junctions. In such a net the spread following a single stimulus is often bounded and does not reach the periphery of the colony. Since repetitive stimulation usually increases the responding area (Horridge, 1957; Josephson, 1961a), the spread following a single shock does not reach all points in the net. The most probable explanation for this incomplete spread is that the excitation stopped when it reached junctions which were non-transmissive. This evidence indicates that the junctions in a coelenterate nerve net are often a mixed population of transmissive and non-transmissive types, at least to the first-arriving impulse.

As stated above, the area excited by a single stimulus in a colonial coelenterate can usually be increased by repeating the stimulus one or more times. This is taken to indicate that some of the junctions of the living nerve net which are not transmissive at the first stimulus become transmissive because of the activity evoked by the stimulus. This process is called facilitation. A second stimulus must follow the first stimulus within some critical time interval for it to increase the spread of excitation, indicating that the state of facilitation disappears or decays with time in these animals.

In the simulation model the junctions between conducting elements are of two types, T and F. The T-junctions are transmissive at all times. An F-junction, on the other hand, is normally non-transmissive. But if an impulse reaches an F-junction, that junction becomes temporarily transmissive for succeeding impulses as a result of "facilitation". Thus the set of transmissive junctions in a network at any particular moment is composed of all T-junctions plus those F-junctions which are temporarily in a transmissive state due to facilitation. In those experiments where only one stimulus is applied to the network, some F-junctions may become transmissive, but since there are no succeeding stimuli, this facilitation can have no effect on the spread of excitation. It is assumed that the

refractory period of conducting elements is sufficiently great to prevent reverberation cycles following a single stimulus.

The introduction of two distinct types of junction, T and F, is admittedly artificial, but it improves the experimenter's control of network parameters without weakening the model. The ratio of T-junctions to F-junctions is controlled, and is usually described here as the "proportion of T-junctions". The actual spatial distribution of T and F junctions over the network is randomly determined in every case.

Facilitated F-junctions are transmissive for some time following the arrival of an impulse. This effect is produced by lowering the threshold of the junction. The transitory nature of this facilitation is simulated by allowing this decrease in threshold to decay linearly with time until the threshold of the junction is so high that it is again non-transmissive. If the time intervals between successive stimuli and the facilitation decay rate at a junction are such that facilitation at the junction decays considerably but not completely between the first two stimuli, several preceding stimuli are required before transmission at the junction is achieved. It should be pointed out that facilitation in the model is equivalent to summation of sub-threshold events until the threshold is surpassed.

The effectiveness of a second stimulus in increasing the spread of excitation in a living nerve net often depends upon the interval between it and the first stimulus—the longer the interval, the less the increment of spread (Horridge 1957; Josephson 1961a). One way to explain this would be to assume that facilitation decays at different rates at different junctions. An impulse arriving a short time after a previous impulse would find many normally non-transmissive junctions facilitated and transmissive; an impulse arriving a longer time after a previous impulse would find fewer facilitated junctions.

In the simulation model there are differences in the facilitation decay rates at different F-junctions. Each F-junction falls into one of four classes with respect to the rate at which induced facilitation decays. The relative proportion of junctions in each of these four classes can be specified, but the spatial distribution of members of each class is randomly determined. The actual rate of facilitation decay for each of the four classes is another variable controlled by the investigator.

The type and properties of each of the two junctions at each crossing can be assigned independently or, if the experimenter so wishes, can be made identical. There is as yet no data from living systems to indicate which is more probable, similar or dissimilar junctions at a neuron crossing. If the experimenter chooses independently determined junctions for an experiment with the model, transmission from horizontal to vertical may require facilitating impulses while transmission in the opposite direction

does not. If both are F-junctions, their facilitation decay rates may be different.

Yet another controlled variable concerns the directionality of facilitation. An impulse arriving at a crossing can facilitate any F-junctions the crossing might have, or only facilitate an F-junction which is part of the conducting element along which the impulse arrives. In the former case the facilitation is bi-directional; in the latter, uni-directional. The model was constructed so that the experimenter can easily specify which mode of facilitation will be used in an experiment. It was hoped that by using these two modes of facilitation, experimentation on the model could show the manner of stimulation which would best reveal differences in the spread of excitation when facilitation is bi-directional and when it is uni-directional. Such information could be compared with experiments on actual nerve nets and would, perhaps, help to determine whether facilitation is uni-directional or bi-directional in the living systems.

C. METHOD OF SIMULATING SPREAD OF EXCITATION

The method of simulating the spread of excitation in the model depends, in part, on the fact that in the net all crossings are the same distance from their nearest neighbors. It is assumed that all elements in the net have equal conduction velocities. Thus the time for conduction of impulses between any two successive crossings on an element is a constant. If the excitation starts at a number of crossings simultaneously, it will spread to the next group of crossings at a uniform rate, and will synchronously arrive at all those crossings after a fixed time interval, Δt. If it is further assumed that transmission across a junction is instantaneous, the excitation will continue to spread everywhere in cadence, with all impulses arriving at crossings simultaneously. The spread of excitation can be followed by having the computer determine, Δt by Δt, just which crossings are reached by impulses.

Provision was made in the model to allow simultaneous stimulation of many elements since, as Horridge (1957) pointed out, the stimulating electrodes used in experiments on coelenterates are many times larger than the neurons, and many nerve cells are probably simultaneously activated. The number and position of stimulated elements in the model can be varied and experiments performed to test the effects of such changes.

A unique number is assigned to each crossing in the model net, its "address". To stimulate a crossing, its address is placed on a list, the entries of which represent all the crossings initially activated. Impulses may be thought of as spreading out along both of the conducting elements perpendicular at each of these crossings. From this first list, another list is made of all junctions which will receive impulses after the time interval

Δt. When all of these junctions are so listed, time in the simulation experiment advances by Δt, and the newly formed list is used to create yet another list, that of the junctions which will receive impulses after another Δt. One list of excited junctions is used to form another, repeatedly, until no new junctions become excited; that is, until all impulses in the net have died out. If the experimenter wishes, the net is then stimulated again, and the entire process is repeated.

Since the refractory period of elements in a nerve net is usually quite long, at least 50 milliseconds in most cases measured (Pantin, 1935a; Bullock, 1943; Nicol, 1955; Horridge, 1958), it is assumed for simplicity in processing impulses that any point on a conducting element can be activated only once during the excitation produced by a single stimulus; it will be refractory to impulses other than the first one impinging on it during this time.

When an impulse arrives at a T-junction, it is transmitted to the perpendicular element, providing that the element is not refractory. And, of course, an impulse continues along the original element until it reaches the end or meets another impulse coming from the opposite direction.

In the case of an F-junction, it is necessary to examine the degree of facilitation to find out whether an arriving impulse will be transmitted to the perpendicular element. Associated with each F-junction is a state-variable S, which represents the degree of facilitation, and a threshold constant θ. An F-junction is transmissive whenever $S \leqslant \theta$. In the simulation experiments described here, θ had the value 0·5, and S always had an initial value of 2·0. Facilitation is simulated by decreasing S by 1·0 each time an impulse arrives at the junction. If, as a result, S is driven below θ, then the impulse is transmitted across the junction (providing again that the perpendicular element is not refractory), and S is set equal to θ. Decay of facilitation is simulated by causing S to increase as a linear function of time (the decay rate) until it reaches a maximum of 2·0 or is reduced again by another arriving impulse.

With these parameter values, an impulse will be transmitted only if $S \leqslant 1{\cdot}5$ when it arrives at an F-junction. At least one preceding facilitating impulse is required to reduce S to this range. The effect of facilitation lasts a period of time determined by the decay rate. A junction which is activated is reached by the excitation some time (= number of Δt's) after the stimulus is delivered to the net. This time depends chiefly on the distance of the junction from the stimulated area. If the stimulated area is not changed between successive stimulations, the time interval between the initiation of activity and its arrival at a particular junction will be approximately constant. Therefore the time interval between successive impulses arriving at an F-junction is about the same as the time interval between the

stimuli which initiated the activity, and the latter can be used in determining the amount by which facilitation has decayed.

This method of simulating facilitation, admittedly artificial, gives the desired results: junctions which are not transmissive can be made transmissive by previous activity; the effects of previous activity decrease with time; and the rate of this decrease is different at different junctions.

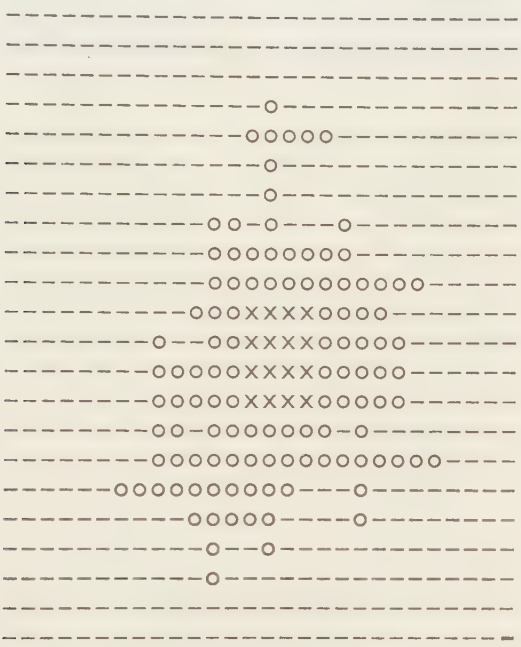


FIG. 3. Part of a two-dimensional print-out of excitation spread. Each character represents one crossing. –, an unexcited crossing; o, a crossing reached by the activity; x, one of the crossings initially stimulated. This particular example shows the spread following a single stimulus to a net with an average element length of 4. One fourth of the functions are of type T.

The most easily measured parameters of excitation-spread in coelenterates are the size of the responding area and the distance of spread. These two measures, as well as a graphic, two-dimensional print-out (see Fig. 3) were chosen as outputs of the simulation model. The area of spread is determined by counting the number of crossings reached by the excitation, and a measure of the distance of spread is obtained by taking the square root of this number. Since it is more easily visualized, the distance of spread is the preferred measure in the following discussion. Because the spread is rarely symmetrical, the square root of the area is only an approximate measure of distance of spread.

3. Experiments with the Model

A. FACTORS AFFECTING SPREAD FOLLOWING A SINGLE STIMULUS

i. The proportion of transmissive junctions.

The following experiment was performed to test the effects of changing the proportion of T-junctions. Sixteen crossings forming a small square in the middle of a net were stimulated and the resulting spread measured. Following each such trial, the proportion of T-junctions was changed and the net stimulated again until a series of proportions, ranging from no

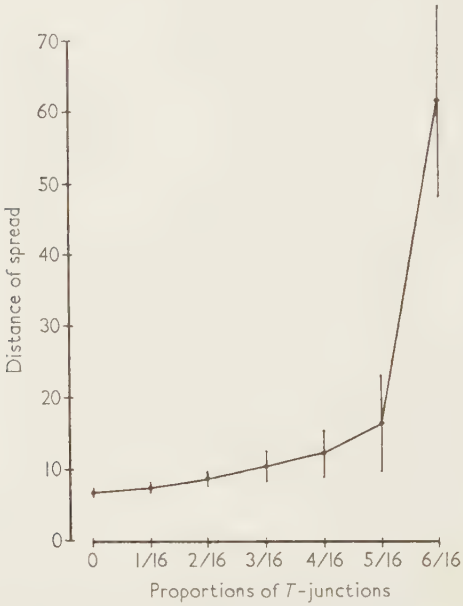


FIG. 4. Spread of excitation following a single stimulus as a function of the proportion of T-junctions. Each point is the average of 20 determinations. The vertical bars (only half-shown for one point) give the standard deviation. The nets used had average element lengths of 4 and the properties of the junctions at each crossing were independently determined.

T-junctions to 6/16 T-junctions, had been tested. Following each series, a new net was created and another series of trials made. No two measurements with a particular proportion of T-junctions were made using the same net. The nets used in this experiment were composed of conducting elements averaging four units long.

Measurements were made on nets in which the properties of the two junctions at each crossing were assigned independently and on nets in which the properties of the two junctions at each crossing were made equal. Twenty measurements were made for each proportion of T-junc-

tions when using nets with independently determined junctions, and 10 trials were made for each proportion in nets where both junctions at each crossing have the same properties. The curves obtained for each case were almost identical. Whether the junctions at a crossing are similar or dissimilar does not seem to affect spread following a single stimulus.

Figure 4 gives the results of the experiments with independently determined junctions at the crossings. It can be seen that the distance of spread

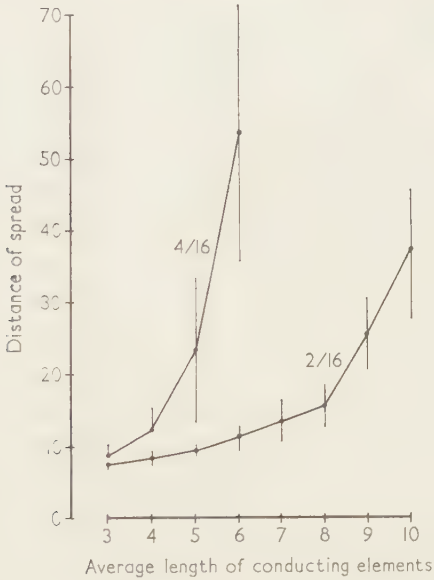


FIG. 5. Spread of excitation following a single stimulus as a function of the average element length. The number above each curve is the proportion of T-junctions for that curve. Each point is the average of 10 determinations. The standard deviations are shown by the vertical bars (only half-shown in some cases). The properties of the junctions at each crossing were independently determined.

plotted against the proportion of T-junctions gives a curve with a positive, increasing slope. The distance of spread with 6/16 T-junctions should actually be somewhat greater than shown. In 19 out of the 20 trials, a portion of the perimeter of the responding area reached the edge of the net.

Unlike spread with a lower proportion of T-junctions, spread in nets with 6/16 T-junctions often deviates from a circular pattern, and frequently long, projecting arms of excited crossings are seen in the graphic print-outs.

A few trials were made on nets with 7/16 T-junctions. Such nets are through-conducting and excitation usually reaches all edges of the net. The spread is not ubiquitous even in these cases, for there are still spaces representing unexcited elements scattered throughout the responding area.

These results show that small changes in the proportion of T-junctions can change a net from one which responds locally to one which is through-conducting. Also interesting is the increased absolute and relative variability of the response with increasing proportions of T-junctions as it is measured by the standard deviation of the distance of spread. This is in agreement with Horridge's similar finding from experiments with his mechanical model.

ii. The average length of conducting elements

Since increasing the lengths of elements in the net increases the number of crossings per element, one expects the average element length to be an important factor influencing the distance of spread. The distance of spread following stimulation of a square of 16 crossings is shown in Fig. 5, plotted as a function of the average element length and for two proportions of T-junctions. Each point represents the average of 10 determinations; each determination was made on a different net. The properties of the junctions at each crossing were determined independently. Although the average element length was varied, the frequency distribution of element lengths about the mean was constant, and was always similar to the distribution shown in Fig. 2.

The distance of spread when plotted against the average length of conducting elements gives a curve of increasing slope, just as it does when plotted against the proportion of T-junctions. Increasing the length of the elements is also similar to increasing the proportion of T-junctions in that the variability of the spread increases and the spread becomes more irregular, tending to deviate from a circular pattern. Since increasing the element lengths and increasing the proportion of T-junctions both increase the probability of activity in one element exciting another element, it is not surprising that they have similar effects on the distance of spread.

iii. Number and density of stimulated crossings

Horridge (1957) proposed that the number of neurons initially activated is important in determining the distance of spread in a nerve net. The effect of changing the number of crossings stimulated, therefore, was an important factor to test in this simulation study.

It must be pointed out that the number of crossings stimulated is usually not equal to the number of conducting elements activated. Stimulating one crossing initiates impulses in two elements, both those intersecting at the crossing. Stimulation of two crossings lying on the same element would initiate activity in fewer elements than would stimulation of two crossings not sharing a common element. Increasing the number of

crossings stimulated, nevertheless, does increase the number of conducting elements activated.

A small square containing 100 crossings was designated as the stimulus area. The locations of the crossings in this square were listed in random order, using a random number table. The first entry on this list was stimulated and the resulting spread measured. Then the first five entries were stimulated, the first 10, the first 20, and so on until all 100 crossings of the stimulus area were stimulated. In this way both the number of crossings stimulated and the density of stimulated crossings was increased.

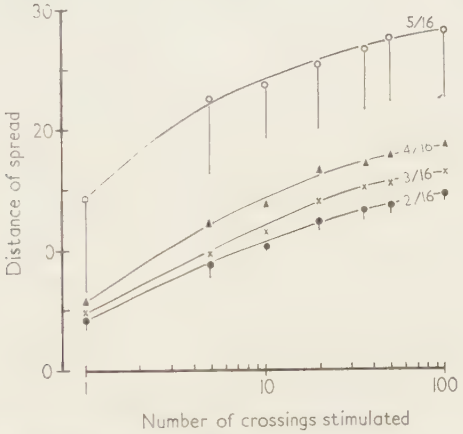


FIG. 6. The spread of excitation following a single stimulus as a function of the number of crossings stimulated. Measurements were made on nets with four different proportions of T-junctions (indicated by the number above each curve). Each point is the average of 20 determinations. The distance to one standard deviation below a point is shown by the vertical bars for two of the curves. For further explanation see text.

Stimulating sets of randomly chosen crossings within the stimulus area makes the simulation experiments more analogous to biological experiments by avoiding systematic effects of a particular spacing of stimulated crossings. The nets used were composed of elements averaging four units long. The properties of the junctions at each crossing were determined independently.

Following seven trials, each with a larger number of crossings stimulated, a new net was created and another seven trials made. A new, randomly-ordered list of crossings in the stimulus area was used after each five sets of trials. The results of 20 sets of trials for each of four proportions of T-junctions are shown in Fig. 6.

It can be seen that increasing the number of crossings stimulated increases the distance of spread but with an effectiveness which decreases as the number of crossings stimulated increases. The distance of spread

plotted against the log of the number of crossings stimulated gives a curve with decreasing slope.

The effectiveness of increasing the number of crossings stimulated in increasing the distance of spread depends on the proportion of T-junctions. Increasing the number of stimulated crossings from 1 to 100 increased the distance of spread 270% in nets with 2/16 T-junctions and only 97% in nets with 5/16 T-junctions.

As Horridge (1957) postulated, the variability of spread decreases as more units are initially activated. This decrease is not impressively great, however. Taking the curve with 5/16 T-junctions as an example, the standard deviation in the distance of spread declined from 7·7 to 5·6 with a hundredfold increase in the number of crossings stimulated. Expressed as a percentage of the distance of spread, the standard deviation declined somewhat more strikingly, from 53% to 20% with an increase from 1 to 100 crossings stimulated.

In the previous experiment, the density of stimulated crossings was varied by changing the number of crossings stimulated in a constant area. The following experiment was performed to test the effect of changing the density of stimulated crossings by varying the dispersion of a constant number of stimulated crossings.

Sixteen crossings were randomly selected from within squares of crossings of increasing size, ranging from 5 to 17 units on a side. Since these crossings were selected from areas of different size, stimulating them initiated activity from different densities of stimulated crossings.

After a series of trials had been made, each with a different density of stimulated crossings, a new net was created and the procedure was repeated. New groups of 16 crossings were randomly selected after every five sets of trials. In all, 20 sets of trials were run with each of two proportions of T-junctions. The nets used had elements averaging four units long and the properties of the junctions at each crossing were independently determined.

The results of this experiment are shown in Fig. 7. The points for the stimulated crossing densities of 1 (16 crossings stimulated in a stimulus area of 16) were obtained from the experiment on the effects of changing proportions of T-junctions.

There are often unexcited areas scattered throughout the responding area when a number of widely scattered crossings are stimulated. At low densities of stimulated crossings, therefore, the square root of the number of crossings reached ceases to be a good measure of the distance of spread. Because of this, the results of this experiment are expressed as the number of crossings reached rather than the distance of spread.

With a constant number of stimulated crossings, fewer crossings are

reached by the excitation as the density of the crossings stimulated is increased. The following explanation is offered for this and for the decreasing slope of the curve of distance versus number of crossings stimulated obtained in the previous experiment.

As the density of stimulated crossings in an area is increased, the area becomes "saturated". There are fewer and fewer unexcited elements in the area which could be activated. Since stimulating one crossing often initiates activity in a number of elements, a number depending in part on the average element length and proportion of T-junctions, complete saturation of the stimulus area can be expected well before all the crossings in the area are stimulated and half-saturation before half the crossings are

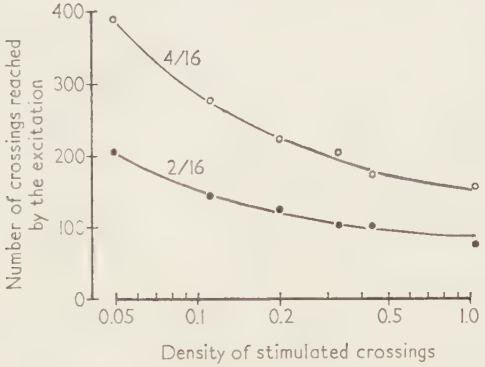


FIG. 7. The number of crossings reached by excitation following a single stimulus as a function of the density of stimulated crossings. The number of crossings stimulated was constant (16). The number above each curve is the proportion of T-junctions for that curve. Further explanation in text.

stimulated. With larger proportions of T-junctions and further spread from each crossing stimulated, saturation can be expected sooner than with smaller proportions of T-junctions. This explains the differences in the effectiveness of the number of crossings stimulated with different proportions of T-junctions. Saturation of an area leads to interaction between active elements by occlusion of possible pathways, and can be expected to be important in living nerve nets as well as in the simulation model.

B. THE EFFECTS OF FACILITATION

i. Spread following repetitive stimulation

Interneural facilitation is probably one of the most important factors affecting spread in coelenterates, and it was primarily to test the effects of this parameter that the simulation model was constructed.

Time becomes a factor in the model with repetitive stimulation—the time between stimuli and the time it takes facilitation at an F-junction to decay. The units of time used in the model are arbitrary, and are not useful for comparison with experiments on living animals since a given interval of time can be expected to have different effects on the decay of facilitation in different species of coelenterates and for the same species at different temperatures (see Hall & Pantin, 1937). The important factor is the time between stimuli as related to the decay rates in the system being investigated. For this reason, time coordinates will not be used in considering the results of experiments with repetitive stimulation. Rather the results will be considered in terms of the proportion of F-junctions in the net which would, if reached by the excitation created by one stimulus, be facilitated and transmissive at the time the next stimulus is delivered.

In the experiments on repetitive stimulation, 16 crossings in the center of a net were given a number of stimuli separated by equal time intervals. The computer program allowed up to seven successive stimuli. The facilitation decay rates and decay rate distributions were adjusted so that the desired proportion of F-junctions would remain facilitated over the time interval separating the stimuli. The properties of the two junctions at each crossing were determined independently and facilitation in these experiments was uni-directional.

Following one trial involving a number of stimuli, the proportion of F-junctions which remained facilitated was changed, and another trial made with the same net. This procedure was repeated until five different proportions had been tested with the net. A new net was created after each set of trials, and the proportions tested again.

Measurements were made on four types of nets, differing as to the average lengths of conducting elements or the proportion of F-junctions. Ten trials were made for each proportion of F-junctions remaining facilitated with each type of net used, except those with average element lengths of four and with 4/16 T-junctions. In the experiments on these nets, 15 trials were made for each proportion.

The results of these experiments are shown in Figs. 8 to 11. In Fig. 8 the results are shown both as the average distance of spread following each successive stimulus and as the increment in the distance of spread due to each stimulus. In the other figures the results are shown as increments of spread.

The number of stimuli given to a net depended upon the type of net and the proportion of F-junctions remaining facilitated. When a high proportion of F-junctions remained facilitated, the excitation often reached the edge of the net by or before the seventh stimulus. The spread is artificially restricted when it reaches the edge of the net, and the potential effectiveness

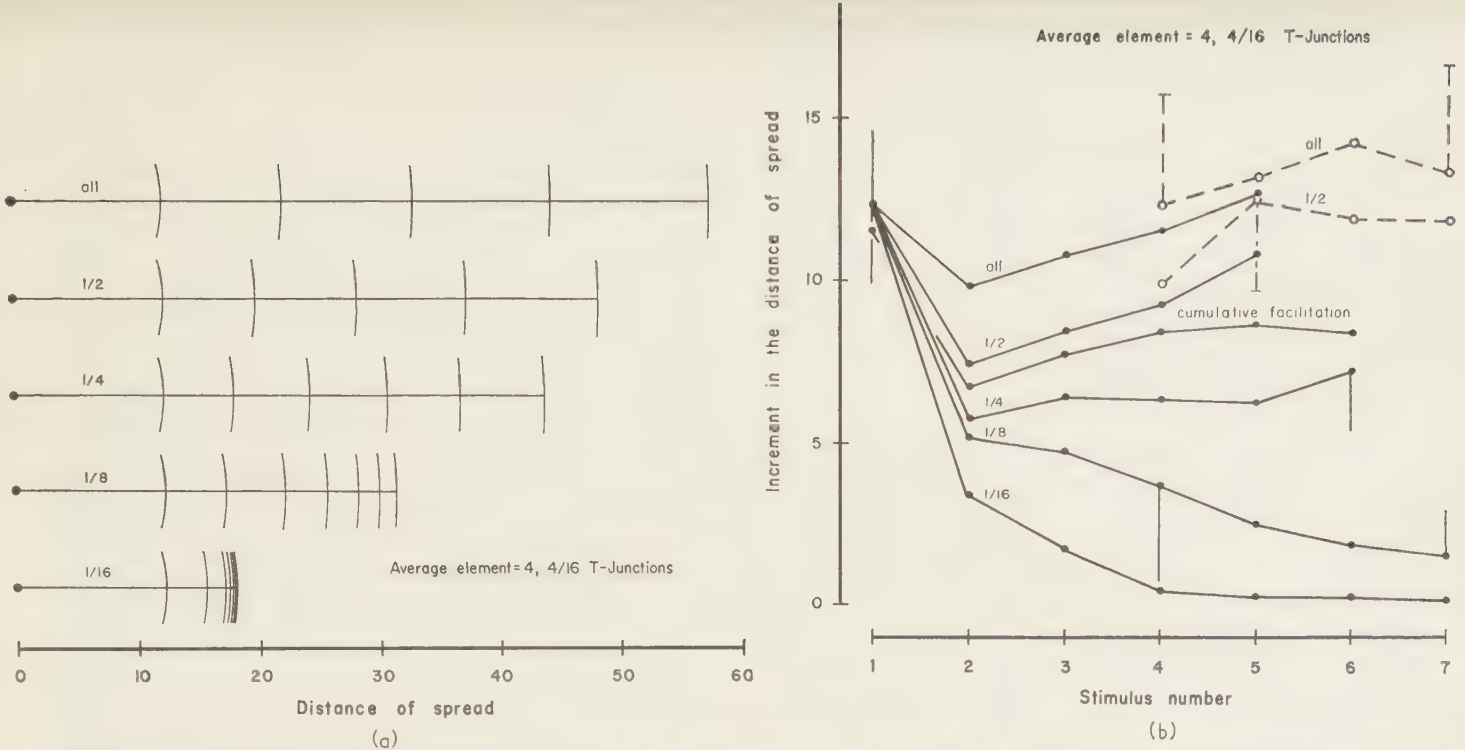


FIG. 8. Spread following repetitive stimulation in nets with average element lengths of 4 and with 4/16 T-junctions. Each point is the average of 15 determinations. The fractions refer to the proportion of F-junctions which would, if reached by an impulse, remain facilitated until the arrival of the next impulse. Facilitation was uni-directional and the properties of the junctions at each crossing were independently determined.

(*a*) The total distance of spread following each successive stimulus. The upright lines indicate the radii of concentric circles.

(*b*) The increments in the distance of spread due to each stimulus. The open circles joined by dotted lines are determinations which were made by stimulating in the corners of nets. The vertical bars give the standard deviations for some representative points (only half-shown in most cases). The curve marked "cumulative facilitation" is described in the text.

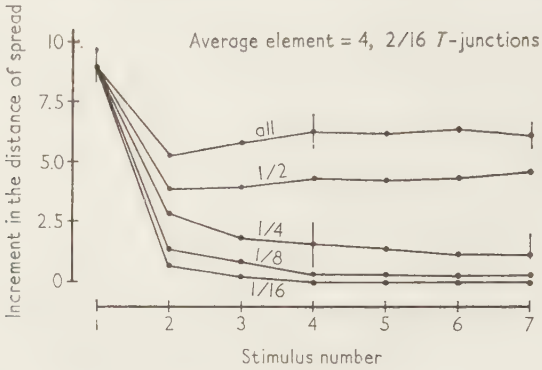


FIG. 9. The increments in the distance of spread due to each stimulus of a series to nets with average element lengths of 4 and with 2/16 T-junctions. Each point is the average of 10 determinations.

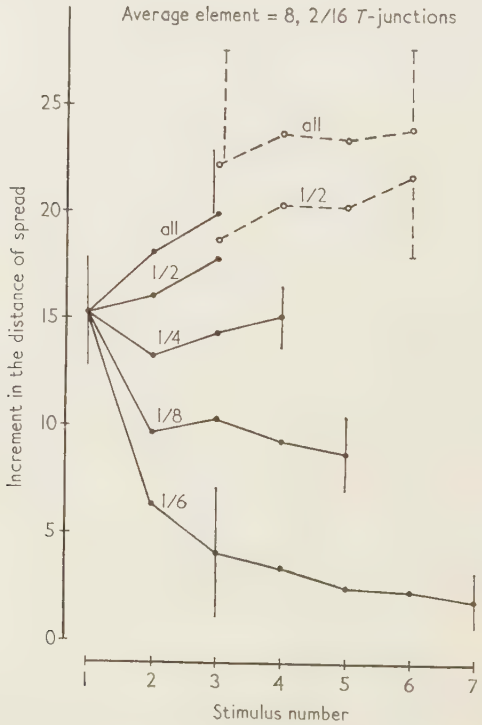


FIG. 10. The increments in the distance of spread due to each stimulus of a series to nets with average element lengths of 8 and with 2/16 T-junctions. Each point is the average of 10 determinations.

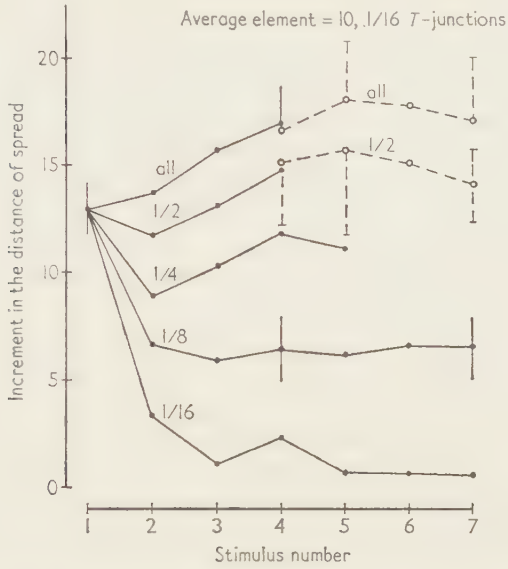


FIG. 11. The increments in the distance of spread due to each stimulus in a series to nets with average element lengths of 10 and with 1/16 T-junctions. Each point is the average of 10 determinations.

of a stimulus initiating activity which reaches the edge cannot be evaluated. For this reason, the number of stimuli given for each type of net and proportion of F-junctions which remained facilitated was always small enough to ensure that the excitation would not reach an edge.

In some cases it was desirable to know the effects of more stimuli than could be given to the center of a net without having the excitation reach the edges. In these instances a slightly different procedure was adopted. Sixteen crossings near a corner were stimulated and the resulting spread measured in just one quadrant. This effectively doubled the linear dimensions of the net. The crossings stimulated were near a corner but not directly in a corner. To minimize edge effects a distance equal to the length of at least one conducting element was left between the two edges of the net and the quadrant of interest. The measurements of spread obtained in this way are indicated by the open circles joined by dotted lines. None of the points obtained by stimulation of crossings in the corner of nets were statistically significantly different (5% level) from their counterparts obtained by stimulation of crossings in the center of nets.

The curve marked "cumulative facilitation" deserves special mention. In these trials, the decay rate distributions and facilitation decay rates were adjusted so that 1/4 of the F-junctions reached by an impulse were facilitated and transmissive at the arrival of the next impulse, 1/4 required two

previous impulses, 1/4 three impulses, and 1/4 required four facilitating impulses before they were facilitated sufficiently to be transmissive.

In all four types of nets used, if the proportion of F-junctions remaining facilitated is low, the increments in the distance of spread due to each successive stimulus become progressively smaller. The lower curve of Fig. 9 actually reaches the zero line; in not one of the ten trials resulting in this curve did more than the fourth stimulus cause any increase in the distance of spread. It seems likely that some or all of the other curves characterized by a continuously negative slope would have also reached the zero line if enough stimuli had been given.

The usual response to repetitive stimulation is spread in approximately equal increments. The spread on the first stimulus in a particular net depends, of course, on the number and density of stimulated crossings and is not a function of facilitation. For stimuli other than the first, there may be an initial increase or decrease in the increments, but all the curves except those which continue to approach the zero line appear to reach a more-or-less stable plateau. Some of the curves show slightly increasing or decreasing increments for the last one or two stimuli, but, because of the variability and the small sample size, these changes are not statistically significant.

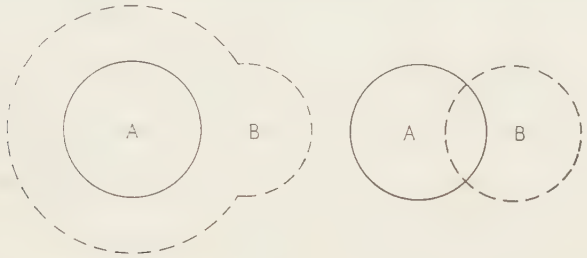
It is of interest to note that conditions in the nets leading to variability in the spread following a single stimulus also lead to variability in the increments of spread due to subsequent stimuli. This is seen in the standard deviations shown for some representative points on the curves.

ii. Uni-directional v. *bi-directional facilitation*

When an impulse arrives at a crossing it can, at the experimenter's discretion, facilitate either *any* F-junctions the crossing might have or only an F-junction *from* the excited element *to* the perpendicular element. In the former case the facilitation is bi-directional, in the latter uni-directional. Absolutely no difference was found in the spread of excitation following repetitive stimulation of a group of crossings when facilitation was uni-directional and when it was bi-directional. This was verified a number of times in nets in which the properties of the two junctions at each crossing were independently determined and in nets in which the properties of the junctions at each crossing were the same. One could not, then, expect to distinguish between uni-directional and bi-directional facilitation in coelenterate nerve nets on the basis of spread following repetitive stimulation of the same area.

One other method of stimulation was tried to differentiate between uni-directional and bi-directional facilitation. Two groups of crossings were selected such that there was slight overlap of the areas excited by stimulation of each alone. One of these groups was stimulated and, after a short

time interval, the other was stimulated. The facilitation decay rates were adjusted so that all F-junctions reached by the first stimulus were still facilitated at the time of the second stimulus. Trials were made with both uni-directional and bi-directional facilitation and with independently and non-independently determined junctional properties at each crossing. The results of these experiments are shown diagrammatically in Fig. 12.



Bi-directional facilitation

1. Similar junctions at each crossing
2. Independently determined junctions at each crossing

Uni-directional facilitation
Independently determined junctions at each crossing (usual response)

Uni-directional facilitation

1. Similar junctions at each crossing
2. Independently determined junctions at each crossing (occasional)

FIG. 12. Spread following sequential stimulation of two neighboring groups of crossings. The solid line shows the spread resulting from stimulation of the first group (the crossings at A); the broken line shows the spread resulting from stimulation of the second group (the crossings at B).

If the properties of the junctions at each crossing are independently determined, spread following sequential stimulation of two groups of crossings is usually the same whether facilitation is uni-directional or bi-directional. In both cases the spread initiated by stimulation of the second group of crossings can extend beyond its normal boundary and sweep through and beyond the area which had responded to stimulation of the first group of crossings. This result was always seen with bi-directional facilitation and was the typical response seen with uni-directional facilitation, although in some cases with uni-directional facilitation the excitation spread no further than it would have if the first group of crossings had not been stimulated.

When the junctions at each crossing are similar and facilitation bi-directional, spread following stimulation of the second group of crossings was always like the usual response with independently determined junc-

tions at a crossing, the excitation extends through and beyond the area responding to stimulation of the first group of crossings. If the facilitation is uni-directional, however, spread following stimulation of the second group of crossings is the same as if the first group had not been stimulated.

If a similar experiment were performed on a coelenterate nerve net and it was found that spread following stimulation of one area was independent of stimulation of an adjacent area, there would be good reason to suspect that there were two reciprocally polarized junctions with *similar* properties at each neuron-crossing or, functionally equivalent, a single non-polarized junction; and that facilitation was uni-directional. Unfortunately the converse result, spread extended because of stimulation of an adjacent area, leaves two possibilities: (*a*) two, sometimes dissimilar junctions at each neuron crossing or (*b*) bi-directional facilitation of the interneural synapses. No method of stimulating the model has been found which permits one to distinguish between these possibilities by observing the spread of excitation.

4. Discussion

A. EXCITATION SPREAD AND REPETITIVE STIMULATION

Horridge (1957) described three types of excitation spread in different coral species following repetitive electrical stimulation: spread in equal distance increments following each successive shock, spread in increasing increments, and spread in decreasing increments. Two of these three types of response, spread in equal increments and spread in decreasing increments, were seen in the simulation model following repetitive stimulation of a group of crossings, and, in fact, were the only responses seen with repetitive stimulation after the first few stimuli.

Horridge proposed that spread in decreasing increments is a result of an increasing number of units being initially activated by each stimulus of a series. But this response is characteristic of nets in which impulses are started from a *constant* number of conducting elements during repetitive stimulation if few of the non-transmissive junctions in the net become facilitated. Since there is no evidence for an increase in the number of elements initially excited with repetitive stimulation, and since there is a more likely alternative, Horridge's explanation for spread in decreasing increments should be rejected as an unnecessary hypothesis.

On the basis of Pantin's (1935a) concept of interneural facilitation, spread in equal distance increments following repetitive stimulation can be expected from nerve nets in which all the units are joined by non-transmissive junctions, if all the non-transmissive junctions reached by an impulse remain facilitated and transmissive until the arrival of the next impulse. The increments of spread in such a case would be equal to the

average radial extension of a conducting unit. Experiments on the simulation model show that spread in equal increments is also a usual response in nets with mixed populations of T and F-junctions between conducting units. Further, the simulation experiments show that the size of these increments is not necessarily related to the size of the conducting elements. The size of the equal increments of spread can vary considerably in the *same* net with changes in the proportion of F-junctions which would, if reached by an impulse, remain facilitated until the arrival of the next impulse.

Spread in continually increasing increments was not seen in the simulation model, and its explanation presents a problem. Horridge's mathematical model did not give such spread until he introduced a new factor, spatial facilitation, to account for it. This property would mean that as more elements in an area are activated, proportionately more neighboring elements become excited. It is difficult to see how spatial summation could operate in the more familiar coelenterate nerve nets with two and only two nerve fibers meeting at each crossing. During spread in such nets there is but a single pre-synaptic fiber at each synaptic area. The special case of a net with multiple pre-synaptic fibers at many of the junctions would be required if spatial facilitation is to be an important factor in the spread of excitation.

Horridge found spread in increasing increments in the polyp retraction responses of two alcyonarian species. Parker (1920) gave evidence that the same nerve net controls both polyp retraction and luminescence in another alcyonarian, *Renilla*. Nicol (1955), studying luminescence in the latter species, found that repeated stimulation often engendered extended periods of activity, indicating spontaneous repetitive firing in the nerve net. This repetitive firing was first seen only after a number of stimuli had been given (the number ranged from a few shocks to "prolonged stimulation" in different specimens), and the duration of the repetitive activity was correlated with the number of previous stimuli. It is suggested, as an alternative to Horridge's proposal, that spread in increasing increments is due to such repetitive activity in the nerve net, an after-discharge which becomes more pronounced the larger the number of stimuli given.

The fact that spread following repetitive stimulation often occurs in equal increments has bearing on the question of the effect of the number of crossings stimulated. The increment of spread due to any stimulus after the first may be viewed as equivalent to the spread outward following *stimulation* of all the crossings on the periphery of the previously responding area which have junctions remaining facilitated. If all the non-transmissive junctions reached by a stimulus remain facilitated, the number of crossings on the periphery "stimulated" by the next stimulus is proportional to the

circumference of the previously responding area and increases as the number of stimuli given the net increases. Yet the increments of spread do not continually increase, but reach a plateau. Increasing the number of crossings stimulated, therefore, need not lead to greater distance of spread.

B. RELIABILITY OF CONDUCTION

It seems likely that there would be an evolutionary advantage to having reliable conducting systems, systems which respond with little variability. It has been shown from experiments on the model that a nerve net like that of coelenterates has little variability in its responses only when the conducting elements have few connections or when the proportion of transmissive junctions is low. But such nets sacrifice distance of conduction for reliability. The factors which lead to greater variability also lead to more elements becoming involved in the spread and hence spread which is greater relative to the lengths of the individual elements.

One way to increase the distance of spread without losing reliability would be to fire a train of impulses following stimulation, that is, to increase spread by repetitive firing. This may be shown by some examples. About the same average distance of spread is attained following one stimulus in nets with average element lengths of 4 and with 5/16 T-junctions (16·3) as following three stimuli to nets with average element lengths of 4, 2/16 T-junctions, and 1/2 of the F-junctions reached by an impulse remaining facilitated (16·9). But the standard deviation in the distance of spread is 6·7 in the first case and only 1·6 in the second. If all of the F-junctions remain facilitated, the average spread is about the same following six stimuli to nets with average element lengths of 4 and 2/16 T-junctions (38·9) as it is following one stimulus to nets with average element lengths of 10 and 2/16 T-junctions (37·2). Again the dissimilarity in the standard deviations is striking, 2·4 for the former, 9·7 for the latter. It is interesting in this respect that repetitive firing is such a pronounced response in coelenterates following a mechanical stimulus, presumably a type of stimulation which the animals would normally encounter (see, for example, Moore, 1926; Pantin, 1935b; Passano and Pantin, 1955; Josephson, 1961b).

C. GENERALITY OF THE RESULTS

The results of the experiments on the simulation model are probably applicable to some types of two-dimensional, diffuse nets other than those composed solely of bi-polar elements. The considerations leading to this conclusion are best explained on the basis of a net with similar junctions at each crossing, but they also apply to nets with independently determined junctions at each crossing.

A net with similar junctions at each crossing and with a moderate proportion of T-junctions can be viewed in several ways. It can be viewed as a net of linear elements joined by some T- and some F-junctions as has been done in the previous discussion. But since a number of elements all joined by transmissive junctions could be considered a single functional element, the net could also be regarded as a net made up of irregularly shaped elements compounded from linear elements; some still linear, some "T" shaped, some "+" shaped, and many with more complex shapes, all joined by facilitatable *non-transmissive* junctions. Or the net could be viewed as being composed of elements somewhat simpler than the last but still not all linear, with the elements joined by some transmissive and some facilitatable non-transmissive junctions. No matter how the net is viewed, the spread is still the same. It is therefore concluded that the types of spread found in the model would also be characteristic of some other types of nets, nets composed of more complex elements.

We wish to thank Dr. T. H. Bullock for offering much valuable advice during the course of this work, and for reading the final manuscript. We would also like to thank Drs. E. J. Batham, G. A. Horridge, and G. O. Mackie for permission to use portions of their figures (in text Fig. 1); and to thank the Director and the personnel at the Western Data Processing Center, University of California, Los Angeles, for their kind assistance. This study was carried out while one of us (R. K. J.) held a National Science Foundation pre-doctoral fellowship.

REFERENCES

Bethe, A. (1903). "Allgemeine Anatomie und Physiologie des Nervensystems". Thieme, Leipzig.
Bozler, E. (1927). *Z. Zellforsch.* **5,** 244.
Bullock, T. H. (1943). *J. cell. comp. Physiol.* **22,** 251.
Hall, D. M. & Pantin, C. F. A. (1937). *J. exp. Biol.* **14,** 71.
Horridge, G. A. (1954a). *J. exp. Biol.* **31,** 594.
Horridge, G. A. (1954b). *Quart. J. micr. Sci.* **95,** 85.
Horridge, G. A. (1956). *J. exp. Biol.* **33,** 619.
Horridge, G. A. (1957). *Phil. Trans.* **B240,** 495.
Horridge, G. A. (1958). *J. exp. Biol.* **35,** 369.
Josephson, R. K. (1961a). *J. exp. Biol.* (in press).
Josephson, R. K. (1961b). *J. exp. Biol.* (in press).
Mackie, G. O. (1960). *Quart. J. micr. Sci.* **101,** 119.
Moore, A. R. (1926). *Amer. J. Physiol.* **76,** 112.
Nicol, J. A. C. (1955). *J. exp. Biol.* **32,** 619.
Pantin, C. F. A. (1935a). *J. exp. Biol.* **12,** 119.
Pantin, C. F. A. (1935b). *J. exp. Biol.* **12,** 156.
Pantin, C. F. A. (1952). *Proc. roy. Soc.* **B140,** 147.
Parker, G. H. (1919). "The Elementary Nervous System". Lippincott, Philadelphia.
Parker, G. H. (1920). *J. exp. Zool.* **31,** 475.
Passano, L. M. & Pantin, C. F. A. (1955). *Proc. roy. Soc.* **B143,** 226.
Passano, L. M. (1958). *Anat. Rec.* **132,** 486.
Passano, L. M. & McCullough, C. B. (1960). *Anat. Rec.* **137,** 387.
Rand, H. W. (1909). *J. exp. Zool.* **7,** 189.
Yamashita, T. (1957). *Z. Biol.* **109,** 116.

J. Theoret. Biol. (1961) **1**, 488–497

The Clone-Size Distribution of Mutants Arising from a Steady-State Pool of Vegetative Phage†

CHARLEY STEINBERG ‡ AND FRANK STAHL §

Division of Biology, California Institute of Technology, Pasadena, California, U.S.A.

(*Received* 6 *June* 1961)

Calculations are presented giving the distribution of clone-sizes of mutants arising in a "steady-state pool" of vegetative phage. It is assumed that replication and maturation are random-in-time processes. The equilibrium pool model is an idealization of current notions about a major portion of the phage life cycle. The calculations are compatible with experimental measurements of the mutant clone-size distribution. It is also shown that the model can account for the observed variance in total burst size.

1. Introduction

In 1951, Luria studied the frequency distribution of the number of spontaneous phage mutants arising in single cell bursts. His results were in very satisfactory agreement with the following model: one intracellular phage "gene" gives rise to two genes, each of which in turn gives rise to two genes, and so on. This we will refer to as exponential growth. He further assumed that the replication process is completely synchronized so that all lines of descent are strictly equal. It is easily shown that such a regime leads to the probability, y_n, that a mutant clone consists of n individuals, which is given by

$$y_n = 1/n, \text{ if } n \text{ is an integral power of } 2,$$

$$= 0, \text{ otherwise.}$$

But replication is not completely synchronized and clone sizes other than integral powers of two are observed. Luria circumvented this

† Taken from a Ph.D. thesis submitted to the California Institute of Technology by C. Steinberg. Aided by a research grant from the National Foundation.

‡ National Science Foundation predoctoral fellow, during part of the course of this work. Present address: Genetics Institute, University of Cologne, Cologne, Germany.

§ National Science Foundation, Division of Medical Sciences, postdoctoral research fellow during part of the course of this work. Present address: Institute of Molecular Biology, University of Oregon, Eugene, Oregon.

difficulty by using the accumulated probability Y_n, that a clone contains *n or more* mutants, which for integral powers of 2 is given by

$$Y_n = 2/n$$

A quantitative comparison of experiment with theory is then made by plotting log Y_n against log n, the prediction for an exponential growth process being a linear relationship with slope -1. Luria's scheme is, of course, just an idealized version of the growth habit of bacteria, for which the theoretical mutant distribution has been greatly elaborated by many authors since the problem was introduced by Luria and Delbrück (1943). (For later references see Armitage, 1953.)

The currently accepted notion of phage growth differs somewhat from the above (see Hershey, 1958). A mature phage is an inert particle, which, in particular, shows no measurable rate of mutation or genetic recombination. After adsorption to a sensitive host cell, the genetic determinants of the phage are injected into the cell. The phage is then in the so-called vegetative state; it is non-infectious but endowed with the ability to replicate and undergo genetic recombination. After some delay, the number of vegetative phages increases exponentially for a time. Then vegetative phage begin to be removed from the "pool" and matured. Such mature phage are inert, like their parents, and do not participate further in the growth processes of the vegetative phage pool. The rate of maturation balances the rate of replication, resulting in a steady-state pool of vegetative phage which is of constant size.

Thus it is seen that current notions about phage growth involve exponential replication, but with markedly unequal lines of descent. It might be intuitively felt, perhaps, that formally this is not too different from the bacterial scheme. Furthermore, the good agreement of the experimental results with the very simple theory did not prompt the consideration of more complicated alternatives. So the matter stood until 1956 when F. W. Stahl and G. Streisinger definitely confirmed the tentative conclusion of Stahl (1956) that the distribution of rare phage recombinants in single cell bursts is quite different from the distribution of mutants. According to ideas current at that time, there was no reason why rare recombinational events should give rise to a different clone-size distribution than rare mutational events. Therefore, Stahl and Streisinger wondered whether their results might not be what is to be expected from steady-state pool kinetics (which would, of course, then call for a special explanation of Luria's results). This was the motivation for the present study. We have written the solution for the case of a vegetative pool of statistically constant size where maturation and multiplication are concurrent random-in-time processes. Levinthal (in a personal communica-

tion in 1956) has considered a similar case. He supposes, for convenience, that the pool undergoes alternating doublings by multiplication and halvings by maturation. Our analysis is probably more realistic (see Visconti and Delbrück, 1953), certainly more elegant, and, in any event, emphasizes the effect of unequal lineages to the greatest degree possible.

2. Mathematical Exposition

Let us first describe the concepts introduced in the preceding section in a somewhat more explicit and precise fashion. We are concerned with a collection of elements or "particles" which may be classified in two ways: by "genotype" and by "state". With respect to genotype, a particle may be "wild-type" or "mutant"; with respect to state, a particle may be "vegetative" or "mature". Thus, there are four distinguishable types of particles. The state of the system as a whole is specified by the number of particles of each type. A vegetative particle of a given genotype may undergo three types of transitions—"replication", "maturation", and "mutation". Replication is a transition from one vegetative particle of a given genotype to two particles of the given genotype. Maturation is a transition from the vegetative to the mature state without change of genotype. Mutation is a transition from one genotype to the other without change of state. A mature particle undergoes no transitions whatsoever. In these terms, our problem is the following: given N vegetative, wild-type particles at time zero, what is the probability, $\bar{P}_n(T)$, that there are exactly n mature, mutant particles at time T? An explicit solution of this problem requires further specifications concerning the transition probabilities.

We postulate that replication and maturation are independent Poisson processes (see Feller, 1957) with identical rate constants. That is, if $p(t)$ is the probability that a particle replicates (or matures) in time t, then

$$\lim_{t \to 0} \frac{p(t)}{t} = k \tag{1}$$

where k is a constant independent of the time of the last replication. Since the parameter, k, is the same for maturation and replication, it is evident that the average number of vegetative particles is a constant, N. For convenience, we will choose a time scale such that $k = 1$.

Concerning the mutational transition, the only biologically interesting cases are those in which the probability is extremely small. Accordingly, we will restrict ourselves to the limiting distribution. Specifically, we assume that one and only one mutational transition occurs in the time interval $(0, T)$. Furthermore, since the average number of vegetative

particles is constant, it follows that the time of the mutational transition is uniformly distributed over (0, T).

Let $P_n(x)$ = the probability that a mutant, vegetative particle present at time zero will give rise to n *mature* mutant particles at time x.

$q_n(x)$ = the probability that a mutant, vegetative particle present at time zero will give rise to n *vegetative* mutant particles at time x.

First we consider a time interval (0, h) and assume, for the moment, that the mutation has occurred at time zero. When h is small, we need consider only the following three cases: the mutant particle matures (probability q_0); the mutant particle replicates once (probability q_2); the mutant particle neither replicates nor matures (probability q_1). It follows from equation 1 that the probabilities for more complex events (involving more than one transition) will be second order or higher in h and would vanish in the limit we shall presently take. It also follows directly from equation 1 (and recalling the convention that $k = 1$) that the relevant probabilities, correct to the first order in h, are:

$$\begin{aligned} q_0(h) &= h \\ q_1(h) &= 1-2h \\ q_2(h) &= h \end{aligned} \tag{2}$$

We now consider an adjacent time interval (h, $t + h$) where t is not necessarily small. At time h there are 0, 1, or 2 mutant vegetative particles. Since these are mutually exclusive possibilities, we can write an expression for $P_n(t + h)$ as a sum of three terms: each term consists of a q, say q_r, multiplied by the conditional probability that given r *vegetative* mutant particles at time h, there will be n *mature* particles at time $t + h$. The conditional probabilities can be expressed in terms of the $P_n(t)$ as follows:

(a) *Coefficient of* $q_0(h)$. If there are no mutant vegetative particles at time h, the one present at time zero must have matured. Since mature particles undergo no transitions, there will be one and only one mature mutant particle at any later time. Hence the coefficient of q_0 will be zero in all $P_n(t + h)$ except $P_1(t + h)$, in which case it will be unity.

(b) *Coefficient of* $q_1(h)$. If there is one mutant vegetative particle at the beginning of the interval (h, $t + h$) (there can be no mature mutant particles present), the conditional probability that n mature mutant particles will be present at the end of this interval is by definition $P_n(t)$.

(c) *Coefficient of* $q_2(h)$. If there are two mutant vegetative particles at the beginning of the interval (h, $t + h$), then at the end of this time interval one particle will have given rise to, for instance, i mature mutant particles and the other will have given rise to j mature mutant particles. The conditional probability of this event is the product $P_i(t)P_j(t)$. The conditional probability that there will be exactly n mature

mutant particles at the end of the interval $(h, t + h)$ is the sum of all such products for which $i + j = n$. Accordingly we may write:

$$\begin{aligned}
P_0(t + h) &= q_1(h)P_0(t) + q_2(h)P_0^2(t) \\
P_1(t + h) &= q_0(h) + q_1(h)P_1(t) + q_2(h)[P_0(t)P_1(t) + P_1(t)P_0(t)] \\
P_2(t + h) &= q_1(h)P_2(t) + q_2(h)[P_0(t)P_2(t) + P_1^2(t) + P_2(t)P_0(t)] \\
&\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\
P_n(t + h) &= q_1(h)P_n(t) + q_2(t)[P^*P]_n \cdot \; n \neq 1
\end{aligned} \tag{3}$$

where $[P^*P]_n$ is the nth component of the convolution of the distribution of P_n with itself:

$$[P^*P]_n = P_0P_n + P_1P_{n-1} + P_2P_{n-2} + \ldots + P_{n-1}P_1 + P_nP_0$$

Substituting (2) into (3), rearranging, and dividing by h, we get in the limit $h \to 0$:

$$\begin{aligned}
P_0' &= -2P_0 + P_0^2 \\
P_1' &= 1 - 2P_1 + [P_0P_1 + P_1P_0] \\
P_2' &= -2P_2 + [P_0P_2 + P_1^2 + P_2P_0] \\
&\quad . \quad . \quad . \\
P_n' &= -2P_n + [P^*P]_n, \;\; n \neq 1
\end{aligned} \tag{4}$$

where $$P_n' = \frac{dP_n(t)}{dt}$$

In principle, the problem is now solved. The differential equations (4) could be solved successively for the P_n's. The labor involved, however, leads us to seek an easier method. Therefore, we define the generating function of the P_n as

$$F(s, t) = P_0 + P_1s + P_2s^2 + \ldots = \sum_{n=0}^{\infty} P_ns^n \tag{5}$$

Squaring F and collecting the coefficients of each power of s, we have that

$$F^2(s, t) = [P^*P]_0 + [P^*P]_1s + [P^*P]_2s^2 + \ldots = \sum_{n=0}^{\infty} [P^*P]_ns^n \tag{6}$$

If we multiply the nth equation (4) by s^n and then sum over all the equations we obtain

$$\sum_{n=0}^{\infty} P_n's^n = s - 2\sum_{n=0}^{\infty} P_ns^n + \sum_{n=0}^{\infty} [P^*P]_ns^n \tag{7}$$

which is equivalent to

$$F' = s - 2F + F^2 \tag{8}$$

where $$F' = \frac{\partial F(s, t)}{\partial t}$$

The solution of this differential equation, with the initial condition that $P_0 = 1$ at $t = 0$, is

$$F = 1 - (1 - s)^{1/2} \tanh (1 - s)^{1/2} t \tag{9}$$

Thus far we have assumed that the mutational transition occurred at time $t = 0$. To find the clone-size probabilities for the case that the probability of mutation is uniformly distributed over the interval from $t = 0$ to $t = T$, we average over this time interval, giving equal weight to all times. That is, the desired probabilities are

$$\bar{P}_n = \frac{1}{T} \int_0^T P_n(t)\, dt \tag{10}$$

and the corresponding generating function is

$$G(s, T) = \frac{1}{T} \int_0^T F(s, t)\, dt = 1 - \frac{1}{T} \log_e \cosh (1 - s)^{1/2} T \tag{11}$$

since

$$G = \bar{P}_0 + \bar{P}_1 s + \bar{P}_2 s^2 + \ldots + \bar{P}_n s^n + \ldots \tag{12}$$

Equation 11 is the generating function, in closed form, for the desired probabilities $\bar{P}_n$. We can recover a particular $\bar{P}_n$ by extracting the coefficient of s^n by the usual procedure for a power series expansion. That is, we evaluate the nth derivative of G with respect to s at $s = 0$ and divide by $n!$:

$$\bar{P}_n = \frac{1}{n!} \left(\frac{\partial^n G}{\partial s^n} \right)_{s = 0} \tag{13}$$

These derivatives, which are easily taken for small values of n, become prohibitively difficult for n greater than about 4. Fortunately, it is easy to express the derivatives as an infinite series suitable for numerical calculation. Cosh x may be expanded as an infinite product (see Bromwich, 1926):

$$\cosh x = \prod_{i = 0}^{\infty} \left[1 + \frac{x^2}{(i + 1/2)^2 \pi^2} \right] \tag{14}$$

Therefore we can rewrite equation 11 as

$$G(s, T) = 1 - \frac{1}{T} \sum_{i = 0}^{\infty} \log_e \left[1 + \frac{(1 - s) T^2}{(i + 1/2)^2 \pi^2} \right] \tag{15}$$

Taking the nth derivative, dividing by $n!$, and evaluating at $s = 0$, we arrive at

$$\bar{P}_n = \frac{(2T)^{2n}}{nT} \sum_{i=0}^{\infty} \left[\frac{\pi^2}{(2T)^2 + (2i+1)^2\pi^2} \right]^n \tag{16}$$

This series converges quite rapidly for n greater than 3. For $T \geq 2\pi$, it may be shown that for $n = 3$, 7 terms give an error of less than 0·1%; for $n = 4$, 5 terms give 0·1% accuracy; for $n = 5$–7, 4 terms suffice; while only 3 terms are needed to give the same accuracy when n is 8 or higher.

Equation (16) gives the desired probabilities in terms of a single temporal parameter, T. We can evaluate this parameter independently of the probabilities, $\bar{P}_n$. To do this, we must consider another problem: given N vegetative particles at time zero, what is the probability, $R_B(T)$ that there are exactly B mature particles at time T? That is, we seek the distribution of mature particles, without respect to genotype. We have already solved this problem for the special case of $N = 1$. In this case $R_B(T) = P_n(t)$, with $B = n$, and the generating function, F, is given by equation (9). In the argument leading to equation (3), it was shown that for the case $N = 2$, $R_B = [P^*P]_n$, with $B = n$. It was then shown that $[P^*P]_n$ is generated by the function, F^2. It is easy to generalize this result by induction. We define the generating function of the R_B as

$$H(s, T) = R_0 + R_1 s + R_2 s^2 + \ldots \tag{17}$$

Then

$$H(s, T) = F^N(s, T) \tag{18}$$

where F is given by equation (9). It is, in principle, possible to calculate the entire set of probabilities $R_B(T)$ from equation (18) by methods similar to those used previously. For our purposes, however, it suffices to calculate the mean and variance of the distribution. These moments can be calculated directly from the generating function without explicit evaluation of the separate probabilities. This is achieved by differentiating H with respect to s and evaluating the derivatives at $s = 1$. In particular we have the mean

$$B = \left(\frac{\partial H}{\partial s} \right)_{s=1} \tag{19}$$

and the variance

$$\text{Var}\,(B) = \left(\frac{\partial^2 H}{\partial s^2} \right)_{s=1} - \bar{B}(\bar{B} - 1) \tag{20}$$

In terms of the function, F, these expressions are:

$$\bar{B} = N\left(\frac{\partial F}{\partial s}\right)_{s=1} \tag{21}$$

$$\text{Var}\,(B) = N\left(\frac{\partial^2 F}{\partial s^2}\right)_{s=1} - \frac{\bar{B}^2}{N} + \bar{B} \tag{22}$$

Evaluating the derivatives, we arrive at the following expressions:

$$\bar{B} = N \cdot T \tag{23}$$

$$\text{Var}\,(B) = \bar{B}(2/3T^2 - T + 1) \tag{24}$$

Both $\bar{B}$ and N (or $\bar{N}$, if N is not a constant) are, in principle, measurable quantities, and T may be evaluated from these using equation (23). We shall use the variance relation (equation (24)) in the discussion which follows.

3. Results and Discussion

Our model predicts that a certain proportion of all mutational events will fail to give rise to mature mutant particles. Such events are, of course, not experimentally observable. Accordingly, the calculated probabilities $\bar{P}_n$ must be renormalized by dividing by $1 - \bar{P}_0$ for comparison with experimental data. Conversely, the experimental data would need to be corrected if it were desired to calculate absolute mutation rates. In any event, the correction is small (about 10% for $T = 2\pi$). Koch and Hershey (1959) report that the pool size, N, is "about" 34 in broth media. Burst sizes, $\bar{B}$, are typically about 200 under similar conditions. Neither of these parameters is known precisely and the burst size, at least, varies considerably with different bacterial strains and with different growth conditions. We have assumed the reasonable and convenient value of $T = 2\pi$ for the ratio of burst size to pool size. In Table 1, we give the values of $\bar{P}_n/(1 - \bar{P}_0)$ (for $n = 1 - 20$) calculated on the assumption that $T = 2\pi$.

In Fig. 1 we compare the calculations of Table 1 with the experimental data and theoretical curve of Luria (1951). The agreement is quite reasonable, particularly when it is considered that T was not used as a free parameter to obtain a "best fit" but rather was chosen, if somewhat arbitrarily, on other grounds. It is not surprising that the calculated values fall below the data for large clones, for we have neglected the contribution due to mutations which occur during the eclipse period. This too could be calculated, but only at the expense of adding another parameter. Our main purpose is not to attempt to fit the experimental data precisely; it is rather to show that the steady-state pool concept offers no difficulty in understanding the mutant frequency results.

In Fig. 1 we also present the data of Stahl and Streisinger giving the

TABLE I

The Distribution of Clone Sizes of Mutants Originating in a Steady-State Pool of Vegetative Phage

The values for $n = 0 - 2$ were calculated from the appropriate derivatives of $G(s,T)$ (eq. 11, 13). The values for $n = 4 \cdots 20$ were calculated from the series approximation (eq. 16). The value for $n = 3$ was calculated by both methods; these calculated values differed by less than 0·1%. The parameter T is assumed to be 2π.

Clone size n	Frequency $\bar{P}_n/(1-\bar{P}_0)$	Clone size n	Frequency $\bar{P}_n/(1-\bar{P}_0)$
1	0·562	11	0·00847
2	0·140	12	0·00727
3	0·0702	13	0·00630
4	0·0438	14	0·00549
5	0·0306	15	0·00482
6	0·0229	16	0·00425
7	0·0179	17	0·00376
8	0·0144	18	0·00334
9	0·0119	19	0·00298
10	0·00996	20	0·00266

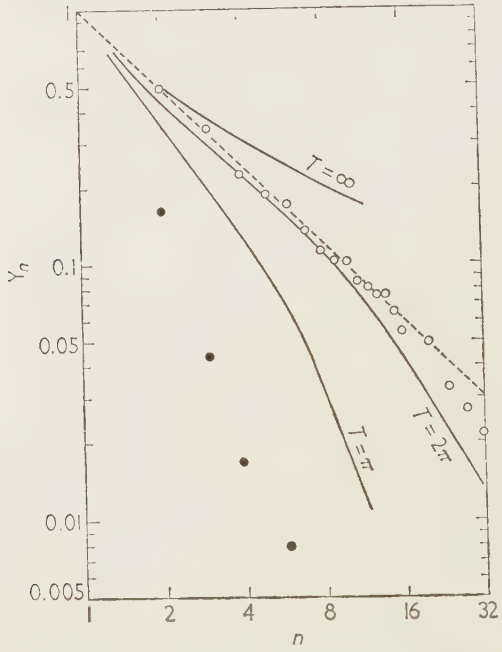


FIG. 1. Luria plots of experimental and theoretical clone-size distributions. The abscissa is the clone size, n; the ordinate is the frequency with which clones of size n *or greater* are found. Both scales are logarithmic. Filled circles are the unpublished data of Stahl and Streisinger (see text) on recombinants for very closely linked markers. Open circles are the data of Luria (1951) on spontaneous mutants. The broken line is the theoretical distribution of Luria. The three unbroken curves are calculated from the steady-state pool model presented here; each curve is labelled with the value of the parameter T (ratio of burst size to pool size) which was used for calculation.

clone-size distribution for recombinants between closely linked markers. We wish to thank Dr. Streisinger for his kind permission to use these unpublished data. These data fall below the calculated distribution, even when T is assumed to be π, which may be considered to be a lower limit. There is now good evidence (Edgar, 1961) that the reason for the disagreement is that recombinant formation is not a simple instantaneous event. It seems likely that the formation of a recombinant for closely linked markers involves first the formation of a non-recombinant heterozygote, which is then followed by the formation of a recombinant heterozygote which in turn may segregate pure recombinants after some delay. The net result is that pure recombinants tend to be formed late in the latent period, and accordingly have little chance to multiply.

An old and repeatedly confirmed observation is that the variance of the phage yields from single cells is large (Delbrück, 1945). This large variance in burst size is undoubtedly due to many causes. For instance, under most conditions of growth, the host bacterial cells are of widely varying sizes. Perhaps, other properties of the cells vary similarly. It should be pointed out, however, that the model for phage replication that we have used here leads to a quite large variance in burst size, even if other sources of variability were absent. We are indebted to Dr. R. H. Epstein for providing the raw data from his "zero-dose" single burst experiment reported elsewhere (Epstein, 1958). In the experiments of Epstein, special conditions were used which eliminate at least part of the usual cellular variability. When these data are corrected for multiple bursts, we obtain a value of 151 for the mean and a value of 7,986 for the variance. This gives a ratio of variance to mean of 19. From equation (24) (and again assuming a value for $T = 2\pi$), we calculate a ratio of about 21. Thus we conclude that a steady-state model in which phage growth and maturation are essentially random processes is sufficient to account for the observed variance in burst size.

REFERENCES

Armitage, P. (1953). *J. Hyg., Camb.* **51,** 162.

Bromwich, T. J. I'a. (1926). "An Introduction to the Theory of Infinite Series", p. 294, 2nd Edition. Macmillan, London.

Delbrück, M. (1945). *J. Bact.* **50,** 131.

Edgar, R. S. (1961). *Virology,* **13,** 1.

Epstein, R. H. (1958). *Virology,* **6,** 382.

Feller, W. (1957). "An Introduction to Probability Theory and its Applications", 2nd Edition. Wiley, New York.

Hershey, A. D. (1958). *Cold Spr. Harb. Symp. quant. Biol.* **23,** 19.

Koch, G. & Hershey, A. D. (1959). *J. mol. Biol.* **1,** 260.

Luria, S. E. (1951). *Cold Spr. Harb. Symp. quant. Biol.* **16,** 463.

Luria, S. E. & Delbrück, M. (1943). *Genetics,* **28,** 491.

Stahl, F. W. (1956). *Virology,* **2,** 206.

Visconti, N. & Delbrück, M. (1953). *Genetics.* **38,** 5.

J. Theoret. Biol. (1961) 1, 498–513

The Precursor–Product Relationship

A. Rescigno† and G. Segre

Istituto di Farmacologia e di Terapia Sperimentale dell'Università, Torino, Italy

(*Received* 8 *February* 1961 *and in revised form* 11 *July* 1961)

This paper is concerned with the relationships between precursors and products in biological systems.

A biological system may be considered as a black box. A transfer function $g(s)$ is defined as the ratio between its output and its input. The output and the input are given by the Laplace transforms of the outflow and of inflow functions. The antitransform of $g(s)$ is the weighting function $G(t)$. If the black box can be decomposed into units, to each unit one transfer function corresponds. The precursor–product relationships are identified with the relationships that exist in the black box between the input and the output or between the input and the output of groups of units.

The function $g(s)$ is an infinitesimal for $s \to \infty$. The order of this infinitesimal permits definition of the *precursor type* and *order*.

In systems of compartments a net may be drawn, in which nodes are the compartments and oriented arms are the transfer functions between them; in these systems the precursor order corresponds to the minimum number of arms in the complete net connecting the precursor to the product.

The sum of the products of the transfer constants of such arms is called the *precursor's principal term* and corresponds to the $(r - 1)$th derivative of $G(t)$ for $t=0$, if r is the precursor order.

Two connected compartments can be classified in several *precursor classes*: complete, unique, absolute, total, partial, with or without recycling.

Applications of these concepts may be found in drug kinetics, in biochemistry and in tracer studies.

1. Introduction

One of the peculiar characteristics of biological systems is the continuous flow of material from point to point and the lasting chemical transformations of substances. The phenomena of diffusion, transport and chemical transformation are therefore of paramount importance. From a methodological point of view, we must then often face the problem of

† Present address: Donner Laboratory, University of California, Berkeley 4, California.

defining the kind of relationship existing between two concentration *v.* time curves in any two points of the system; the two points may eventually coincide in the space (but not in the representative space†), for instance, when we are dealing with the chemical transformation of one substance into another.

This general problem may be considered as the problem of the precursor–product relationship.

A substance in one point may be the precursor of the same substance in another point, when diffusion or transport forces intervene; on the other hand, a substance (or a chemical group or even an atom in a compound) may be the precursor of another substance (or of a chemical group or of an atom in a compound) in the same or in another point of the space when chemical transformations occur, or when chemical transformations are coupled with diffusion or transport.

It is apparent from these statements that the precursor–product relationship displays two aspects, which are experimentally distinct, but which may be treated in the same theoretical manner.

Given two points 1 and 2 in a biological system, it is possible to observe the concentration *v.* time curve of a substance in 1 and in 2. We may ask what kind of relationship exists between the two curves, i.e. we may ask whether the substance in 1 passes directly into 2. Alternatively, if we observe the concentration *v.* time curves of two substances 1 and 2 at the same point, we may ask whether the substance 1 passes directly into the substance 2, through a single chemical reaction, or whether a whole chain of reactions exists between 1 and 2. The former is a problem of transport or of diffusion, the second is a problem of chemical transformation, or of chemical transformation coupled with diffusion or transport, but both are problems of precursors and products. We may even ignore that the transport is directed from point 1 to point 2 or that the chemical transformation is going from substance 1 to substance 2. What kind of evidence can we obtain about the existence of a link between 1 and 2 and about the direction of the transformation, by knowing only the concentration *v.* time curves of the material at 1 and at 2?

The theory of compartments gives a solution to this problem but, in order to apply the compartment theory, several assumptions are necessary, whereas we need a theory with the minimum number of assumptions to be applied also to non-compartmentalized systems. The compartment theory assumes a first-order rate of transfers from one compartment into another, and, first of all, a compartmentalized system; the compartment theory can be applied to the kinetics of foreign substances (drugs), when first-order reactions occur, or to tracer kinetics in steady-state systems,

† In the representative space each variable of the problem represents one dimension.

even when reactions of different order occur (Rescigno, 1954, 1956, 1960; Rescigno & Segre, 1961a).

To what extent it is possible to give an answer to this problem, even without assuming a compartmentalized system? We give a tentative answer in this paper, in which we present a more general approach to the problem. We present also a more general definition of the precursor–product relationship when a system of compartments is dealt with.

2. The Transfer Function

Every biological system may be considered as a "black box" in regard to the kinetics of the material in it; the black box transforms every input into an output.

If $X_1(t)$ indicates the concentration or the amount of a substance 1 (or of a substance at the point 1) and $X_2(t)$ indicates the concentration or the amount of the substance 2 (or of the same substance at the point 2), then the input and the output of the black box are expressed by $x_1(s)$ and $x_2(s)$, the Laplace transforms of $X_1(t)$ and $X_2(t)$.

We may then define the ratio of the two Laplace transforms, under the condition of zero initial material storage, as the *transfer function* $g(s)$ of the black box:

$$x_2(s)/x_1(s) = g(s).$$

If the system is linear, i.e. if the superposition principle can be applied, the transfer function $g(s)$ is a characteristic invariant of the system.

We may point out that all our considerations apply to linear systems only.

When the black box corresponds to a model of a compartment system, it can be decomposed in units, each corresponding to a lag transfer function, i.e. to a function of the form $k/(s + K)$; non-compartmentalized systems can also be sought, and the black box can then be decomposed not only in the above mentioned units, but also in delay units and in units of other types. For each of these units a transfer function may be calculated.

From the mathematical properties of the transfer function $g(s)$ we obtain the properties of the precursor–product relationship between 1 and 2. Because of the superposition principle, the properties of the black box and of $g(s)$ depend upon the nature of the units and on their connections and may be calculated with algebraic rules (Rescigno, 1960; Rescigno & Segre, 1961a), when the units and their connections are known.

The Laplace transform $f(s)$ of a function $F(t)$ is defined as

$$f(s) = \int_0^\infty F(t)\, e^{-st}\, dt.$$

For the existence of $f(s)$ it is necessary that this integral has meaning. The variable s may be real or complex ($s = \sigma + i\omega$).

There are several methods for obtaining with good approximation the Laplace transform $x(s)$ of an experimentally known function $X(t)$; one method is that of Guillemin (1953); graphical methods (Rosenbrock, 1959; Teasdale, 1955) and methods employing mechanical devices (Reynolds, 1955; Gardner, 1957) can also be used.

It may be recalled that, by knowing the functions $X_1(t)$ and $X_2(t)$, it is possible to calculate by numerical approximation the function $G(t) = \mathscr{L}^{-1}\{g(s)\}$, which is called the *weighting function* between 1 and 2 and is the antitransform of $g(s)$. The weighting function corresponds to the response of the black box to the unit impulse function.

In order to calculate $G(t)$ it is necessary to approximate the two curves X_1 and X_2 by a series of rectangles with equal basis Δt and with heights $A_1, A_2, \ldots$; $B_1, B_2, \ldots$, respectively, and such that in each interval their area is equal to the area described by the curve. The A's and the B's are then ordered as in Table 1, and the indicated operations are carried out,

TABLE 1

B_1	B_2	B_3 ...	A_1	A_2	A_3 ...
$= G_1A_1$	$- G_1A_2$	$- G_1A_3$ ...	G_1	G_2	G_3 ...
	$- G_2A_1$	$- G_2A_2$ ...			
		$= G_3A_1$ ...			

as in ordinary division. The series of G's in the table corresponds to the series of rectangles of base Δt which approximate the function $G(t)$.

3. Type and Order of the Precursors

It can be shown (Churchill, 1944, Section 12) that, if $G(t)$ is bounded, then

$$\lim_{s \to \infty} g(s) = 0;$$

$g(s)$ is an infinitesimal for $s \to \infty$, and the behavior of $G(t)$ at $t = 0$ is related to that of $g(s)$ as $s \to \infty$ along the real axis. The order of this infinitesimal can be found by studying what simple function $\phi(s)$ gives

$$\lim_{s \to \infty} \phi(s) . g(s) \begin{cases} \neq 0 \\ \neq \infty \end{cases}$$

Let us take, for example, $g(s) = A/(s + a)$; this transfer function corresponds to a lag transfer function and its antitransform is given by

$G(t) = A\,e^{-at}$; it is clear that, if we multiply $g(s)$ by s, we obtain $s.g(s) = As/(s + a)$; it is easily seen that

$$\lim_{s\to\infty} s.g(s) = A;$$

$$\lim_{t\to 0} G(t) = A.$$

In this case $g(s)$ is of the order of $1/s$. Of the same order are the functions $1/s$ and $s/(s^2 + a^2)$, whose antitransforms are respectively 1 and cos at.

Of the order of $1/s^2$ are the functions

$$1/s^2;\ 1/(s + a)(s + b);\ 1/(s + a)^2;\ 1/(s^2 + a^2);\ (s^2 - a^2)/(s^2 + a^2)^2,$$

whose antitransforms are respectively

$$t;\ \frac{1}{b-a}(e^{-at} - e^{-bt});\ t\,e^{-at};\ \frac{1}{a}\sin at;\ t\cos at.$$

The order of the function $s^{-3/2}.e^{-k/s}$ (whose antitransform is given by $\frac{1}{\sqrt{\pi k}}\sin 2\sqrt{kt}$) is the order of $s^{-3/2}$; of the same order is the function $\frac{1}{\sqrt{s}\,(s - a^2)}$, whose antitransform is given by $\frac{1}{a}\,e^{a^2t}\,\mathrm{erf}\,(a\sqrt{t})$, where

$$\mathrm{erf}\,(x) = \text{error function of } x = \frac{2}{\sqrt{\pi}}\int_0^x e^{-\lambda^2}.d\lambda.$$

The weighting function must be bounded for $t \geqq 0$ and therefore cannot be a function like erf $(k/\sqrt{s})$ whose antitransform is $F(t) = \frac{1}{\pi t}\sin(2k\sqrt{t})$, because $\lim_{t\to 0} G(t) = \infty$.

In order to calculate $\lim_{t\to 0} G(t)$ it is possible to plot the first value of $G(t)$, i.e. G_1, obtained in the above mentioned approximation, with decreasing values of Δt.

The analytical form of $\phi(s)$ indicates the *type of the precursor*.

When $\phi(s) = s^r$, and r is an entire number, the precursor is called *of rational type*. In this case, if $g(s)$ is given as a fraction of two polynomials, then the numerator is of lower degree than the denominator and $G(t)$ is formed by a sum of exponential terms (system of compartments).

When $\phi(s) = s^r$, and r is a fractional number, the precursor is called *of irrational type*.

In both cases r is called the *precursor's order*.

The function $\phi(s)$ may also be given by other functions; the type of the precursor is characterized by their analytical form and it is named in

consequence. If, for instance, $\phi(s) \quad e^{t_1 s}$, this implies a *delay* t_1 in the appearance of the output, given the input origin at $t = 0$, and the precursor is called *of delayed type*. This means that in the black box some unit introduces a delay in the appearance of the output. It may be recalled that to multiply a function $g(s)$ by $e^{-t_1 s}$ is equivalent to translate its antitransform $G(t)$ by t_1, so that $G(t)$ becomes $G(t - t_1)$.

When $\phi(s) = s^r . e^{t_1 s}$, the precursor is of rational or of irrational type, of order r, with delay t_1.

When $\phi(s) = -s\, e^{-s} \ln s$, the precursor may be called *of logarithmic type*.

The limit

$$\lim_{s \to \infty} \phi(s).g(s) = M$$

is called the *precursor's principal term*.

We may therefore give the following definition:

"In a system, 1 is the precursor of 2 when the transfer function $g(s)$ from 1 to 2 is not identically null and $\lim_{s \to \infty} g(s) = 0$; the type of the precursor is named according to the function $\phi(s)$ such that

$$\lim_{s \to \infty} \phi(s).g(s) \begin{cases} \neq 0. \\ \neq \infty \end{cases}$$

The value of the above limit is the precursor's principal term."

It is evident from the definition that the precursor–product relationships have been identified with the relationships that exist in a black box between the input (precursor) and the output (product).

A. PRECURSOR OF RATIONAL TYPE

From a known property of Laplace transforms (Churchill 1944, Section 58), if $\lim_{s \to \infty} s^r.g(s) = M$, where r is an entire positive number and M is different from 0 and from ∞, then

$$G(0) = G'(0) = \ldots = G^{(r-2)}(0) = 0$$

and

$$G^{(r-1)}(0) = M.$$

Therefore we can give the following definition:

"In a system, when the weighting function from 1 to 2 is not identically null, and

$$G(0) = G'(0) = \ldots = G^{(r-2)}(0) = 0,$$

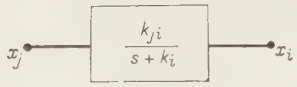
$$G^{(r-1)}(0) = M,$$

where M is different from 0 and from ∞, then 1 is a precursor of 2 of rational type, of order r, and M is the precursor's principal term."

If the first $(r-2)$ derivatives of $G(t)$ are void at $t=0$, but $\lim_{t\to 0} G^{(r-1)}(t) = \infty$, then the precursor is of irrational type.

B. PRECURSORS IN SYSTEMS OF COMPARTMENTS

When the black box is a model of a system of compartments, it can be decomposed into units, each corresponding to a lag transfer function



We may then replace each unit with an *oriented arm* and its input and output with *nodes*. The nodes are connected with all the possible connections. The system of compartments is then represented by a net (Rescigno, 1960), in which the nodes x_i are the Laplace transforms of the amounts (or concentrations) of material in the ith compartment, the oriented arms correspond to the transfer functions between the compartments, $k_{ji} =$ transfer constant from compartment j to compartment i (corresponding to the fraction of material in j that passes in i in unit time) and $K_i =$ sum of the transfer constants from compartment i to all other compartments connected with it.

Because of the superposition principle, the system may then be described by the following system of equations:

$$x_i = \sum_j \frac{k_{ji}}{s+K_i}\, x_j \quad (j = 1, 2, \ldots) \tag{1}$$

From system (1) we obtain

$$g(s) = \frac{x_2}{x_1} = \frac{p(s)}{q(s)}$$

where $p(s)$ and $q(s)$ are two polynomials, the first being of lower degree in s than the second; then 1 is a precursor of 2 of rational type.

We may therefore say: "In a system of compartments, in order that 1 be a precursor of 2, it is necessary that the transfer function be a fraction of two polynomials, with numerator of lower degree than the denominator."

In this case, according to the Heaviside partial fraction expansion, $g(s)$ may be written as a sum of terms of the type

$$\frac{A_1}{s+\alpha} + \frac{A_2}{(s+\alpha)^2} + \ldots + \frac{A_m}{(s+\alpha)^m}$$

$$\frac{B_1 s + C_1}{(s+\beta)^2+\gamma^2} + \frac{B_2 s + C_2}{[(s+\beta)^2+\gamma^2]^2} + \ldots + \frac{B_n s + C_n}{[(s+\beta)^2+\gamma^2]^n}$$

where $-\alpha$ is a real pole (a zero of $q(s)$) of multiplicity m, and $-\beta \pm i\gamma$ is a complex conjugate pole of multiplicity n. The antitransforms of these functions have the form

$$(a_1 t^{m-1} + a_2 t^{m-2} + \ldots + a_m)\, e^{-\alpha t},$$
$$(b_1 t^{n-1} + b_2 t^{n-2} + \ldots + b_n)\, e^{-\beta t} \cos(\gamma t - \delta)$$

If only real, non-multiple poles $-\alpha$'s occur, then $G(t)$ has the form

$$G(t) = \textstyle\sum_j A_j\, e^{-\alpha_j t}.$$

When the degree of $p(s)$ is $(n - r)$ and the degree of $q(s)$ is n, the precursor's order is r.

C. PRECURSOR CLASSES

System (1) represents a system of compartments; if $x_0(s) = \mathcal{L}\{X_0(t)\}$ corresponds to the input of the system and $x_0(s)$ is connected only to node $x_1(s)$ and the compartments 1 to n have zero initial material storage, then system (1) can be written (Rescigno, 1956):

$$\begin{cases} (s + K_1)x_1 - \sum_{j=2}^{n} k_{j1}x_j = x_0 \\ -\sum_{j=1}^{i-1} k_{ji}x_j + (s + K_i)x_i - \sum_{j=i+1}^{n} k_{ji}x_j = 0 \end{cases}$$

In order to solve this system it is necessary to solve the determinant Δ

$$\Delta = \begin{vmatrix} (s + K_1) & -k_{21} & \ldots & -k_{n1} \\ -k_{12} & (s + K_2) & \ldots & -k_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ -k_{1n} & -k_{2n} & \ldots & (s + K_n) \end{vmatrix}$$

we have therefore

$$\frac{x_i}{x_0} = (-1)^{i+1} \frac{\Delta_{1;i}}{\Delta},$$

where $\Delta_{1;i}$ is the minor of Δ obtained by suppressing 1st row and ith column.

The function $X_i(t)$ is given by the antitransform of $x_i(s)$.

The transfer function $g(s)$ between, for example, compartment 1 and 2, is given by

$$g(s) = \frac{x_2}{x_1} = \frac{-\Delta_{1;2}}{\Delta_{1;1}}.$$

From a theorem proved elsewhere (Rescigno, 1956), it follows that:

$$-\Delta_{1;2} = k_{12}{}^{P}\Delta_{1,2} + \Sigma k_{1i}k_{i2}{}^{P}\Delta_{1,2,i} + \Sigma k_{1i}k_{ij}k_{j2}{}^{P}\Delta_{1,2,i,j} + \ldots$$

where ${}^{P}\Delta_{a,b,\ldots}$ is the principal minor of Δ obtained by suppressing the ath, bth, ... rows and the corresponding columns.

By developing $\Delta_{1;1}$ by the elements of its 1st column (the 2nd column of Δ), there follows

$$\Delta_{1;1} = (s + K_2)\Delta_{1,2;1,2} + k_{23}\Delta_{1,3;1,2} - k_{24}\Delta_{1,4;1,2} + \cdots$$

and, on the basis of the previous quoted theorem (Rescigno, 1956), we may write:

$$\Delta_{1;1} = (s + K_2)^P\Delta_{1,2} - \sum k_{2i}k_{i2}{}^P\Delta_{1,2,i} - \sum k_{2i}k_{ij}k_{j2}{}^P\Delta_{1,2,i,j} - \cdots.$$

Therefore

$$\frac{x_2}{x_1} = \frac{k_{12}{}^P\Delta_{1,2} + \sum k_{1i}k_{i2}{}^P\Delta_{1,2,i} + \cdots}{(s + K_2)^P\Delta_{1,2} - \sum k_{2i}k_{i2}{}^P\Delta_{1,2,i} - \cdots} \tag{2}$$

If

$$\begin{cases} k_{1i}k_{i2} = 0, \\ k_{1i}k_{ij}k_{j2} = 0, \\ \cdots\cdots\cdots, \end{cases} \tag{3}$$

$$\begin{cases} k_{2i}k_{i2} = 0, \\ k_{2i}k_{ij}k_{j2} = 0, \\ \cdots\cdots\cdots, \end{cases} \tag{4}$$

for all i's, j's, ...,
formula (2) becomes

$$g(s) = \frac{x_2}{x_1} = \frac{k_{12}}{s + K_2}.$$

In this case, i.e. when (3) and (4) hold, and $k_{12} \neq 0$, we call 1 the *total precursor* of 2.

It is then easily seen that

$$\mathcal{L}^{-1}\{x_2/x_1\} = \mathcal{L}^{-1}\{k_{12}/(s + K_2)\} = k_{12}e^{-K_2t} = G(t),$$

and

$$G(0) = k_{12}.$$

If (4) but not all the equations (3) hold, then the precursor order r is the minimum number of k's composing a product of type (3) not void, and the principal term is the sum of all products of type (3) formed by r k's.

If one or more of the products (3) are different from zero and $k_{12} \neq 0$

$$g(s) = \frac{x_2}{x_1} = \frac{k_{12}{}^P\Delta_{1,2} + \sum_a k_{1a}k_{a2}{}^P\Delta_{1,2,a} + \cdots}{(s + K_2)^P\Delta_{1,2}} =$$

$$= \frac{k_{12}}{s + K_2} + \frac{\sum_a k_{1a}k_{a2}{}^P\Delta_{1,2,a} + \cdots}{(s + K_2)^P\Delta_{1,2}}$$

the precursor is then of first order, $G(0) = k_{12}$, but 1 is no more the total precursor of 2, in as much as not all the material reaching 2 comes from 1, but in part it comes also from 1 through other compartments. In this case

we call 1 a *partial precursor* of 2, and the term $k_{12}/(s + K_2)$ is the *short-circuit term.*

When one of the products (4), for instance $k_{23}k_{32}$, is different from zero, but (3) are otherwise valid, then

$$g(s) = \frac{x_2}{x_1} = \frac{k_{12}{}^P\Delta_{1,2}}{(s + K_2)^P\Delta_{1,2} - k_{23}k_{32}{}^P\Delta_{1,2,3}} =$$

$$= \frac{k_{12}}{s + K_2} \cdot \frac{1}{1 - \dfrac{k_{23}k_{32}{}^P\Delta_{1,2,3}}{(s + K_2)^P\Delta_{1,2}}} =$$

$$= \frac{k_{12}}{s + K_2} \cdot \frac{1}{1 - T_{22}}.$$

We call the factor $1/(1 - T_{22})$ the *recycling factor,* and the precursor is called *precursor with recycling.*

We can also imagine a *partial precursor with recycling.*

When all the k_{1i}'s except k_{12} are zero, 1 is called the *complete precursor* of 2.

When all the k_{i2}'s except k_{12} are zero, 1 is called the *unique precursor* of 2.

When all the k_{1i}'s except k_{12} are zero, and all the k_{i2}'s except k_{12} are zero, 1 is called the *absolute precursor* of 2.

In all these cases the precursor order is one.

If we represent the system of compartments with a net, as described before, we may draw the following schemes of the *precursor classes* (Fig. 1).

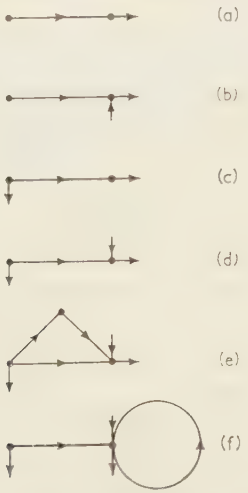


FIG. 1. Precursor classes: (a) absolute precursor; (b) complete precursor; (c) unique precursor; (d) total precursor; (e) partial precursor; (f) total precursor with recycling.

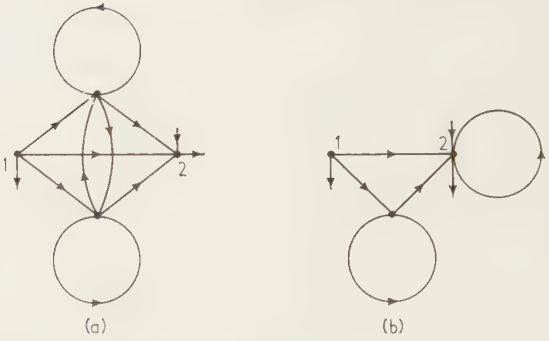


FIG. 2. Precursor classes: (a) partial precursor; (b) partial precursor with recycling.

Other more complicated examples of *partial precursors* and of *precursors with recycling* are given in Fig. 2.

Therefore (3) and (4) allow a partition of precursors in classes.

In the case of *absolute*, *unique* and *complete precursor* (*total precursors*), we have

$$g(s) = \frac{x_2}{x_1} = \frac{k_{12}}{s + K_2}, \tag{5}$$

and

$$G(t) = k_{12}\, e^{-K_2 t}.$$

The calculation of $G(t)$, the weighting function between 1 and 2, gives immediate insight into the type, the order and the class of the precursor.

Zilversmit's rule and the transfer function between two connected compartments

Let X_1 and X_2 indicate the activities of two connected compartments in a system, and V_1 and V_2 their volumes; we may write

$$dX_2/dt = k_{12}V_1X_1/V_1 - K_2V_2X_2/V_2,$$

if the active material comes from compartment 1; if moreover the system is in the steady-state, then

$$\sum_i k_{i2}V_i = K_2V_2.$$

When 1 is a *unique* or an *absolute precursor* of 2 (see the preceeding paragraph), it follows that

$$k_{12}V_1 = K_2V_2,$$

then

$$dX_2/dt = 0 \text{ for } X_1/V_1 = X_2/V_2.$$

This relation is known as Zilversmit's rule (Zilversmit, Entenman & Fishler, 1943): "The curves representing the specific activity X_1/V_1 and X_2/V_2 cross at the time of the maximum of X_2/V_2, or of X_2."

It is clear that Zilversmit's rule applies to only two of the several possibilities of the precursor product relationship, i.e. to unique and to absolute precursors.

Zilversmit's rule is valid in a system of compartments in the steady-state only if compartment 2 receives the material only from compartment 1, i.e. in the following cases (Fig. 3):

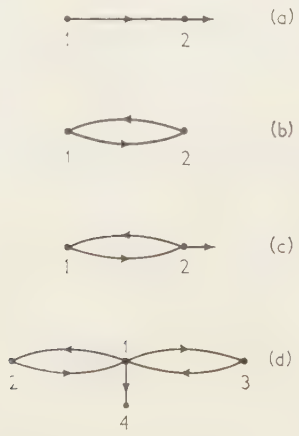
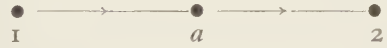


FIG. 3. (a) simply connected, irreversible system; (b) closed system of two compartments; (c) open system of two compartments; (d) mammillary system of four compartments.

(a) simply connected, irreversible systems;
(b) closed systems of two compartments;
(c) open systems of two compartments;
(d) mammillary systems, between the central and the peripheral compartment, and not vice-versa.

All these are cases of first-order precursors.

If the precursor order is >1, Zilversmit's rule does not apply. For instance, in the system



we have

$$g(s) = \frac{k_{1a}k_{a2}}{(s + K_a)(s + K_2)};$$

$$G(t) = \frac{k_{1a}k_{a2}}{K_2 - K_a}\,(e^{-K_a t} - e^{-K_2 t}).$$

If two compartments (a and b) are interposed between 1 and 2, and no short-circuit term is present, then the precursor order is equal to 3 and the principal term is given by $k_{1a}k_{ab}k_{b2}$.

In the case of *complete precursor*

we have

$$k_{12}V_1 < K_2V_2,$$

and therefore, at the time of the maximum of X_2,

$$X_1/V_1 > X_2/V_2.$$

In the case of *precursor with recycling*, the diagram of Fig. 1(f) can be drawn and the transfer function x_2/x_1 will be (Rescigno, 1960; Rescigno & Segre, 1961a)

$$g(s) = \frac{k_{12}}{s + K_2} \cdot \frac{1}{1 - T_{22}}$$

with

$$T_{22} = \frac{k_{23}k_{32}}{(s + K_2)(s + K_3)}$$

then

$$g(s) = \frac{k_{12}(s + K_3)}{(s + K_2)(s + K_3) - k_{23}k_{32}}$$

and the precursor is of the first order. The same order is found in the case of partial precursor (Fig. 1(e)).

It is also clear that it is possible to have *precursors of* n*th order with recycling* (see Fig. 4) and *partial precursors of* n*th order with or without*

FIG. 4. Precursor of order 3 with recycling.

recycling (see Fig. 5). The presence of a recycling factor does not change the order of the precursor; in fact, the recycling factor $1/(1 - T)$ consists of a fraction with numerator and denominator of the same degree in s and therefore its presence does not change the degree difference in s of the transfer function $g(s)$. The same can be said if a recycling is included in the complete diagram between two compartments whose transfer function

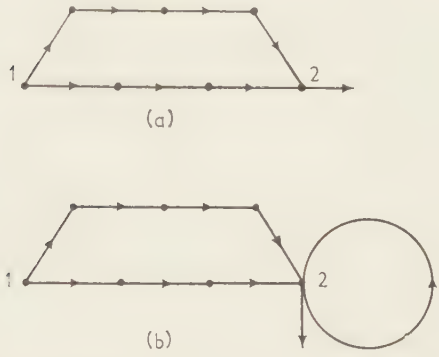


FIG. 5. (a) partial precursor of order 3; (b) partial precursor of order 3 with recycling.

is to be calculated. If in the transfer function a short-circuit term occurs, the precursor order is given by the order of the short-circuit term: this follows directly from the rules of addition of fractions.

By remembering that in the net two tandem arms can be substituted by their product (Rescigno, 1960; Rescigno & Segre, 1961a), it is possible to formulate the following rule: "The precursor order gives the minimum number of arms, in the complete diagram, connecting the precursor to the product."

It follows also that the precursor's principal term M is formed by

$$k_{1a_1}k_{a_1a_2}\ldots k_{a_{r-1}2},$$

the sum being extended to all such products of r k's.

When precursors with delay t_1, of r order, occur, we have

$$\lim_{s\to\infty} s^r . e^{t_1 s} . g(s) = G^{(r-1)}(t_1) = M$$

and M corresponds to the precursor's principal term for the same system but without delay.

D. PRECURSORS IN NON-COMPARTMENTALIZED SYSTEMS

Non-compartmentalized systems can be often found in biology; for instance, in the pulmonary circulation the blood in the right ventricle is obviously the precursor of the blood in the left atrium, but the pulmonary system cannot be conceived as a system of compartments and therefore the pulmonary transfer function cannot be decomposed in lag transfer units; nevertheless it is possible to determine the pulmonary weighting function (Rescigno & Segre, 1961b).

Also in the well-known experiment of Shemin and Rittenberg of determination of erythrocyte life-span using ^{15}N-labelled hemin, the newly

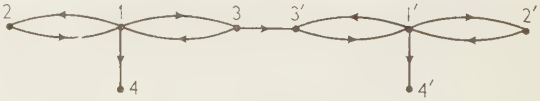
synthesized hemin is the precursor and the hemin outflowed from dead erythrocytes is the product; the corresponding black box can be decomposed in delay units and in units different from lag transfer units, and the weighting function can be easily calculated (Rescigno & Segre, 1961a).

E. PRECURSORS AND DRUG KINETICS

The problem of precursor–product relationship presents itself very often in drug kinetics: when a drug passes from compartment 1 into another compartment 2, the first may be called the precursor of the second one; in an analogous manner, when a substance *a* in some part of the system is transformed into another substance *b*, *a* is called the precursor of *b*, and two connected compartments *a* and *b* must be drawn in the diagram.

Therefore, if we have the following mammillary system

where: 1 = blood; 2 = extravascular space; 3 = liver; 4 = urine, and a substance *a* is irreversibly transformed into the substance *b* in the liver, we can draw the following net



where the numbers 1′, 2′, 3′ and 4′ are referred to the distribution of substance *b* in the same mammillary system where substance *a* is distributed (the parameters with the same number coincide in the space but not in the representative space; see footnote, p. 499).

We see immediately that compartment 3 is the *total precursor* of compartment 3′, whereas compartment 1 is a precursor of 3rd order of compartment 1′. This means that the transfer function $g(s)$ between 3 and 3′ is of 1st order, and $G(0) = k_{33'}$; the transfer function between 1 and 1′, that may be calculated if the concentration *v.* time curves of the two substances in the blood are known, is of 3rd order, and $G(0) = G'(0) = 0$; $G''(0) = k_{13}k_{33'}k_{3'1'}$.

The evaluation of the order of the transfer function between blood and urine, or between blood and bile, allows the analysis and the characterization of the *elimination processes* more thoroughly than the actually employed

methods (Segre, 1961). The precursor order r indicates that $(r \quad 1)$ is the minimum number of compartments connecting in this case the blood to the urine, or to the bile.

The case of *precursors with recycling* occurs in many instances of drug kinetics, and therefore it seems interesting to have a theoretically founded method of its analysis.

Other applications of the precursor-product relationship can be found in biochemistry, when chains of reactions, eventually with recycling, are studied; in the Krebs cycle it has been shown (Reiner, 1953) that Zilversmit's rule applies only to citrate-ketoglutarate reaction; the evaluation of the weighting function between two substances of the cycle, or, at least, the evaluation of $G(0)$, $G'(0)$, ..., would permit an analysis of the precursor–product relationship between any pair of intermediates of the cycle.

A similar analysis could be carried out in the study of photosynthetic reactions, when $^{14}CO_2$ incorporation is followed: if an activity v. time curve could be obtained with sufficient precision for every compound, it would be possible to state whether a compound is the total precursor of another one.

REFERENCES

Churchill, R. V. (1944). "Modern operational Mathematics in Engineering". McGraw-Hill, New York.

Gardner, M. T. (1957). *Control Engng.* **4,** 119.

Guillemin, E. A. (1953). *Proc. nat. Electron. Conf.* **9,** 1954.

Reiner, J. M. (1953). *Arch. Bioch. Biophys.* **46,** 80.

Rescigno, A. (1954). *Bioch. biophys. Acta,* **15,** 340.

Rescigno, A. (1956). *Bioch. biophys. Acta,* **21,** 111.

Rescigno, A. (1960). *Bioch. biophys. Acta,* **37,** 463.

Rescigno, A. & Segre, G. (1961a). "La cinetica dei farmaci e dei traccianti radioattivi". Boringhieri, Torino.

Rescigno, A. & Segre, G. (1961b). *V Congr. Soc. it. Biol. Med. nucleare, Pisa, nov.* Minerva Nucleare. In press.

Reynolds, J. B. Jr. (1955). *Control Engng.* **2,** 60.

Rosenbrock, H. H. (1959). *Control Engng.* **6,** 116.

Segre, G. (1961). *Ann. N.Y. Acad. Sci.,* in press.

Shemin, D. & Rittenberg, D. (1946). *J. biol. Chem.* **166,** 627.

Teasdale, A. R. Jr. (1955). *Control Engng.* **2,** 56.

Zilversmit, D. B., Entenman, C. & Fishler, M. C. (1943). *J. gen. Physiol.* **26,** 325.

J. Theoret. Biol. (1961) 1, 514–528

Proposed Molecular Models: II†. Conformations of Staphylomycin and other Polypeptides and a Possible Relation to the Structure of Water

Donald T. Warner

The Upjohn Company, Kalamazoo, Michigan, U.S.A.

(*Received* 17 *July* 1961)

It is suggested that the conformations of polypeptides in aqueous biological systems may consist of hexagonal arrangements of the component acids or multiple arrangements of these hexagonal units. The peptide bonds may be pictured as occupying positions on one surface of the hexagonal plane ("hydrophilic side") while the amino acid side chains project from the other side of the hexagon ("hydrophobic side"), and the α-carbon atoms lie approximately in the plane. The carbonyl oxygen atoms of the peptide bonds occupy the outer corner positions of the hexagonal units and the α-carbon atoms form a ring within the hexagon, allowing a small central space for "reactive" side chain groups. The molecular models of these conformations show that the polar side chains frequently interact favorably to provide stabilizing hydrogen bonds and electron transfer pathways. Hydrophobic forces from non-polar interactions also may contribute to the conformational stability and provide possible shielding for electron transfer pathways in the proposed models. It is suggested that the carbonyl oxygen pattern of these hexagonal peptide units coincides with the "second neighbor" oxygen pattern of water at physiological temperature ranges to provide an important additional stabilizing force.

1. Introduction

In a previous publication (Warner, 1961) some of the possible tertiary structures for gramicidin S, etamycin, and actinomycin were examined with the aid of molecular models, and conformations that might be preferred in aqueous biological systems were proposed. In general, these proposals suggested three main features:

1. Maximum hydrophobic interactions of the component amino acid side chains, sometimes favored by the presence of D-amino acids.
2. The arrangement of the peptide linkages into hexagonal patterns with the carbonyl oxygen atoms at the corners of the hexagons, and the presence of possible "reactive centers" within the hexagon.

† For the first paper in this series see Warner (1961).

3. The suggested presence of a multiple number of hexagonal units in the more complicated molecules (e.g. two fused hexagons in a cyclic decapeptide).

Because the conformations were presented as probable structures for an aqueous environment, it was especially interesting to find that the distances between the carbonyl oxygen atoms at the hexagonal corners are approximately 4·8 Å, corresponding to the appropriate distance for bonding a water molecule between adjacent carbonyl oxygens. This uniform hexagonal and coplanar arrangement of the six carbonyl oxygen atoms in the cyclic hexapeptide has another implication when compared with the oxygen arrangement of an extended segment of the water structure. If the *primary* or "nearest neighbor" oxygen positions in the water structure are examined, either three or four such "near neighbors" will readily form trigonally or tetragonally bounded planes, respectively, while six "nearest neighbors" form either a boat or chair non-planar hexagonal arrangement. However, when the examination is extended to the *secondary* or "second neighbor" oxygen positions, we find that a group of seven "second neighbor" atoms lies in a definite coplanar hexagonal pattern with six oxygen atoms at the hexagonal corners equidistant from the seventh oxygen atom at the center of the hexagon. In the model of the oxygen positions of the water structure (Plate I), the six "second neighbor" oxygen atoms referred to in this discussion have been outlined with a cord to show the uniform hexagonal arrangement. The concept may also be extended (Plate I) by employing twelve such "second neighbor" oxygen atoms, joining ten to form two fused hexagons, and using the remaining two oxygen atoms for the respective hexagonal centers. All of the twelve "second neighbor" oxygens are definitely coplanar.

Since these "second neighbor" oxygen atoms in the water structure are also approximately 4·8 Å apart, this particular planar segment of the water structure has the same *oxygen* pattern as the *carbonyl oxygen* pattern of a cyclic hexapeptide (or decapeptide) in the conformations which I have proposed. The similarity of shape and planarity are readily seen; but the respective distances in the two structures need further discussion. The liquid water pattern has, of course, a degree of flexibility with temperature, and it must be conceded that there is a particular temperature at which the carbonyl oxygen positions of the cyclic hexapeptide and the "second neighbor" oxygens of the water structure could exactly coincide. I was interested to see whether this temperature might be close to 37°, a frequent temperature for maximum enzyme activity and the average human body temperature. From the data of Morgan & Warren (1938), the values for the "nearest neighbor" oxygen distances in liquid water are known at 1·5°

(2·90 Å) and at 83° (3·05 Å). Since these values are a function of the well-known relative volumes of water at various temperatures (Lange, 1956), a plot was made of the temperatures between 1·5° and 83° against the relative volumes of water at these temperatures. By inserting the values of $(2{\cdot}90)^3$ and $(3{\cdot}05)^3$ on the ordinate at the 1·5° and 83° positions respectively, the interpolated "nearest neighbor" oxygen distance in water at 37° was about 2·93 Å. Using the formula of Morgan & Warren (1938) this corresponds to a calculated "second neighbor" distance of $2{\cdot}93 \times (8/3)^{\frac{1}{2}} = 4{\cdot}79$ Å (37° C).

With regard to the oxygen–oxygen distance in a cyclic peptide, Schmidt, Hodgkin & Oughton (1957) observed in X-ray studies of gramicidin S that a strong reflection occurs at 4·8 Å. No positive assignment of this strong band was made. To obtain the approximate distance between two adjacent

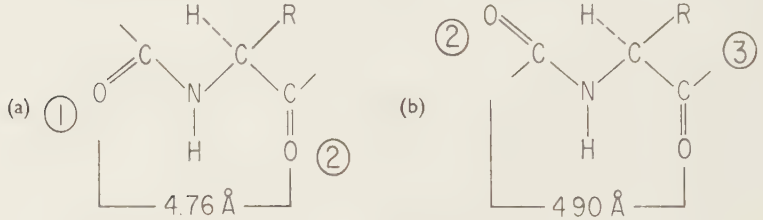


Fig. 1. Carbonyl oxygen distances for *cis* and *trans* peptide bonds.

carbonyl oxygens, I have made careful drawings based on the bond angles and atomic distances for the peptide bond as presented by Corey & Pauling (1953), assuming a planar configuration for all of the atoms. For the *cis* amide bond (Fig. 1(a)), the oxygen–oxygen (1–2) distance is 4·76 Å. A similar drawing of the *trans* amide bond (Fig. 1(b)) gives a value of 4·90 Å. These values are indicative of the approximate magnitude of the oxygen–oxygen distances in planar *cis* or *trans* amide structures although the models indicate that in a cyclic peptide these six atoms probably are not all co-planar. Changes in the distances between the carbonyl oxygen atoms will result from moving some of the atoms slightly out of the plane. In Fig. 1(a) if the oxygen atoms 1 and 2 are both rotated upward through an arc of about 20° around their respective C—N and C—C bonds, (while maintaining the C—N—C—C-atoms coplanar), the oxygen atoms soon reach a separation distance of 4·80 Å. In a similar manner, if oxygen atoms 2 and 3 in Fig. 1(b) are rotated through a similar arc of about 20° above the C—N—C—C plane to the same relative positions, then oxygen atoms 2 and 3 are also 4·80 Å apart. Although these values were only derived from diagrams, it would appear that the oxygen atoms 1, 2, and 3 are approximately the same distance above the C—N—C—C plane at the points

where (1–2) = (2–3) = 4·80 Å. Other atoms by varying slightly from a coplanar position could perhaps produce changes of similar magnitude in the oxygen–oxygen distances. However, our model studies of cyclic hexapeptides seem to indicate that it is the C—O bond which projects upward at an angle of about 20° from the plane defined by the α-carbon atoms. By these slight rotations either the *cis* or the *trans* amide bond (Fig. 1(a) or (b)) in the protein or polypeptide structure has an oxygen–oxygen distance (4·8 Å) which is very close to the "second neighbor" oxygen distance in water at 37° (4·79 Å). From a consideration of the direction of the carbonyl oxygen projections, it seems probable that the *cis* amide linkage would be more favorable for bonding one water molecule between two neighboring amino acid units in the same chain. By the same criterion the presence of *trans* amide units at the junction positions of a

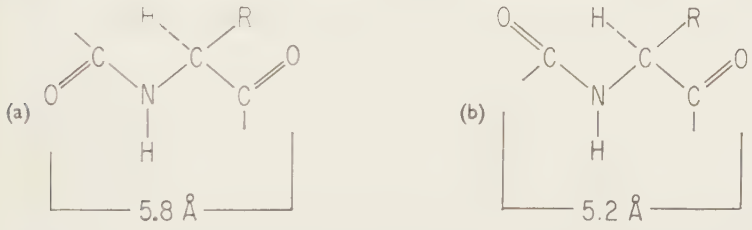


FIG. 2. Carbonyl oxygen distances for alternate *trans* and *cis* peptide bonds.

cyclic decapeptide, for example, could project these centrally located oxygen atoms from opposite sides of the ring into a good position for bonding one mole of water between them. In this sense two carbonyl oxygen atoms from widely separated portions of the peptide structure could be held at a definite space interval by water bonding cooperating with other forces in the molecule (e.g. disulfide linkages).

The peptide oxygen distances in Fig. 2(a) and (b) are about 5·8 Å and 5·2 Å, respectively. Even at 83° the "nearest neighbor" oxygen distance in water is only 3·05 Å, which corresponds to a "second neighbor" value of only 4·98 Å. Consequently water bonding between adjacent amino acid units at 83° or lower would practically be ruled out in these two peptide configurations. Interchain water bonding, however, still would be possible.

It would not be expedient at the present time to carry the conclusions beyond the accuracy of the derived values, which only suggest that the carbonyl oxygen distances indicated in Fig. 1(a) and (b) are very close to the "second neighbor" oxygen distances in water at about 37°. If temperatures considerably above 37° are imposed on the water system, it would be expected that the average dimensions of the hexagonal "second neighbor" oxygen pattern in water will become so much larger than the proposed

hexagonal carbonyl oxygen pattern of the peptide that the latter will begin to lose an orderly contact with its water environment. Such an expansion of the more flexible water pattern beyond the more rigidly bonded peptide (or protein) limits may conceivably be a factor in denaturation phenomena if our suggested conformations are correct. However, since the hydrophobic attraction between non-polar groups may actually increase with increasing temperature (Kauzmann, 1959), complete disruption of the tertiary structure would not necessarily occur as the temperature rises.

Although the close correlation between the oxygen positions in the aqueous environment and the peptide solute may be a plausible argument in support of the hexagonal conformations which I have proposed, the hexagonal concept acquires additional substantiation from other aspects of the tertiary structure of more diverse polypeptide molecules when they are arranged in similar conformations. In the earlier report (Warner, 1961) most of the polypeptides examined with the models had a high proportion of amino acid components with non-polar side chains and secondary amide bonds. I would like to examine one additional example of this type, staphylomycin, because of the interesting orientation of the three aromatic rings which the models suggest. This will be followed by a discussion of a suggested conformation for arginine vasopressin, a polypeptide with more reactive side chains than our previous examples. A third model of the N-terminal decapeptide of ACTH will be presented with some reservations, since probable ester linkages and other bonds are still a matter of conjecture. Some of these conformations will also be discussed from the standpoint of their electron transfer potentialities, especially those resulting from side chain interactions. This approach may furnish significant clues concerning the biological activity and mode of action of these compounds.

2. Staphylomycin

The primary structure of staphylomycin has been elucidated by Vanderhaeghe & Parmentier (1959, 1960), and consists of a cyclic hexapeptide with one lactone linkage incorporating the hydroxyl group of threonine into the main ring while the amino group of threonine is linked to 3-hydroxy-picolinic acid as a branch on this ring. This polypeptide contains several uncommon amino acids such as D-α-amino butyric acid, L-N-methylphenylalanine, and 4-keto-L-pipecolic acid. The position of the various amino acids is indicated by the number code in Plate II (a) and (b). Plate II (a) represents the amide or "hydrophilic" side of the proposed conformation (Warner, 1961), and the hexagonal arrangement of the carbonyl oxygen atoms is clearly seen. Staphylomycin, like etamycin (Sheehan, Zachau & Lawson, 1957) contains a high percentage of secondary amide groups, so that probable —NH—O=C bonds as conformational

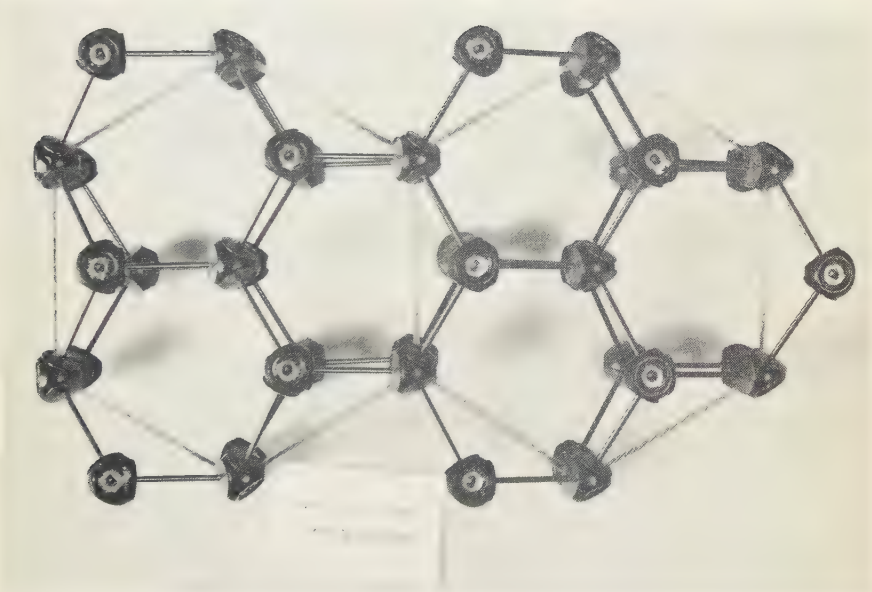
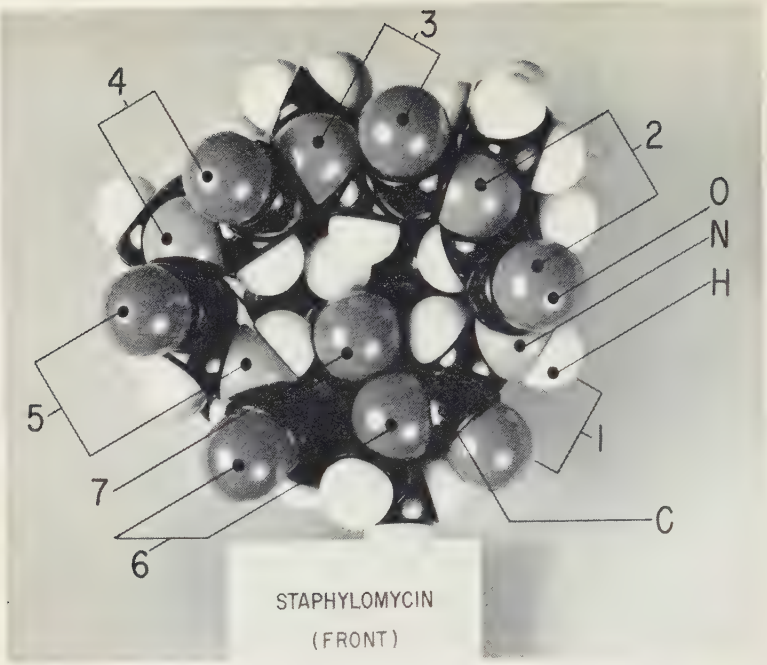


PLATE I. Oxygen pattern of the water structure.

[*To face p.* 518.

(a)

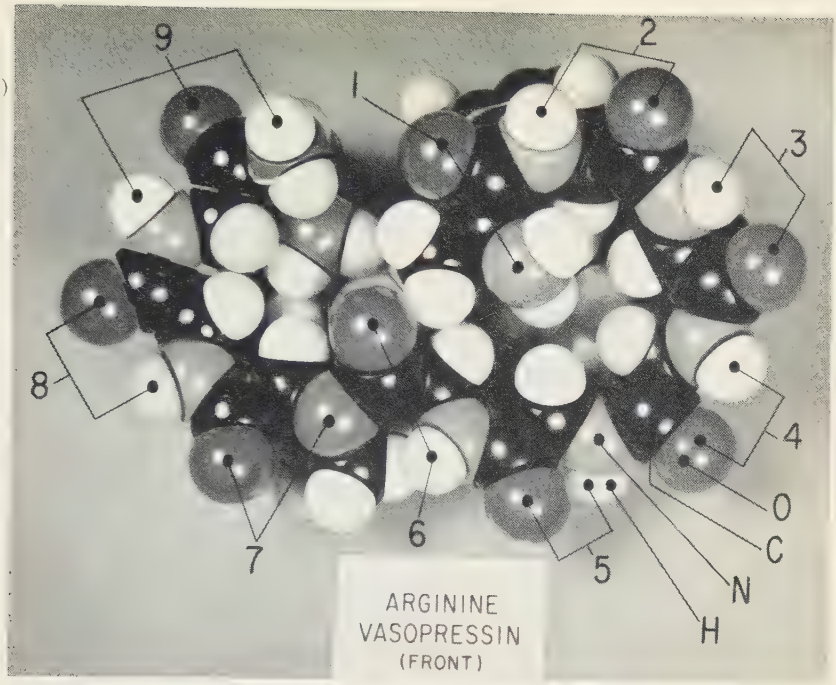


(b)

PLATE II. Staphylomycin, (a) front, (b) back.

1 = L-Phenylglycine
2 = 4-oxo-L-pipecolic acid
3 = L-N-methylphenylalanine
4 = L-proline
5 = D-α-aminobutyric acid
6 = L-threonine
7 = 3-hydroxypicolinic acid

(a)



(b)

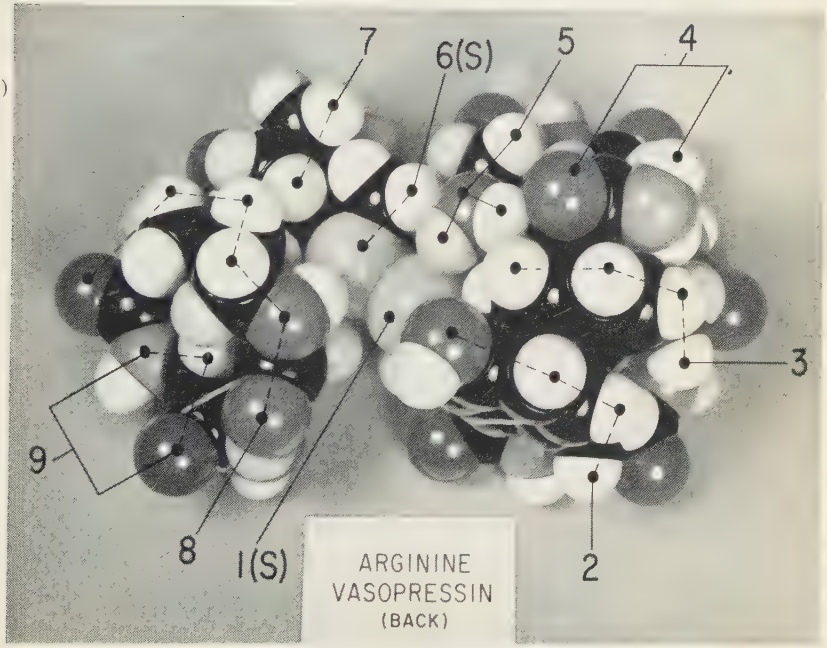
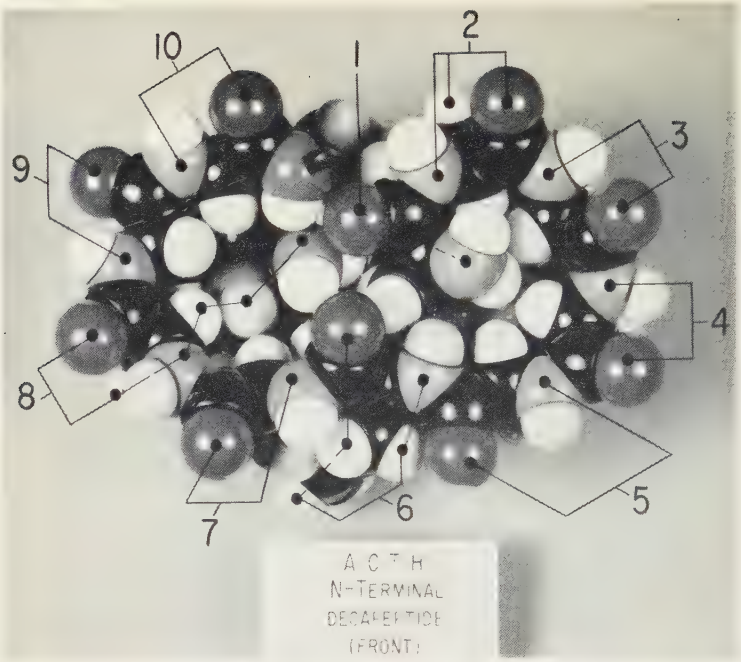


PLATE III. Arginine vasopressin, (a) front, (b) back.

1 = ½-L-Cystine
2 = L-tyrosine
3 = L-phenylalanine
4 = L-glutamine
5 = L-asparagine
6 = ½-L-cystine
7 = L-proline
8 = L-arginine
9 = glycine amide

(a)



(b)

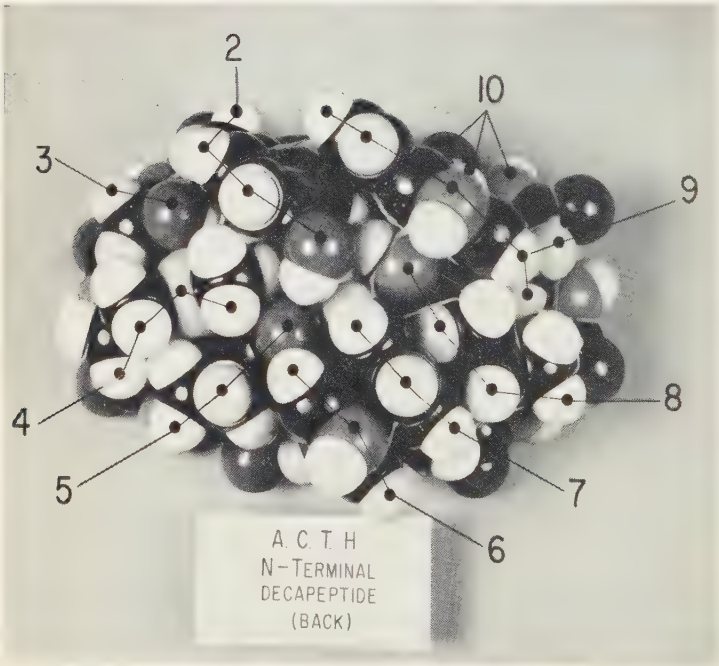


PLATE IV. ACTH, N-terminal decapeptide, (a) front, (b) back.

1 = L-Serine not visible in (b)
2 = L-tyrosine
3 = L-serine
4 = L-methionine
5 = L-glutamic acid
6 = L-hystidine
7 = L-phenylalanine
8 = L-arginine
9 = L-tryptophan
10 = glycine

determinants would be minimized by their absence. The most interesting features of the "front" side of staphylomycin in the suggested conformation are the compact overall shape and the presence of the reactive carbonyl-phenolic hydroxyl of the picolinic acid in the hexagonal center.

The hydrophobic "back side" of staphylomycin (Plate II (b)) in the conformation illustrated here shows an interesting arrangement of the three aromatic nuclei in which the pyridine ring of the 3-hydroxypicolinic acid is sandwiched between the two benzene rings of the phenylglycine and N-methyl-phenylalanine residues. The rubber bands which I have used to simulate hydrophobic forces and also hold the model together tend to bring the rings directly side by side although the slightly staggered arrangement of aromatic rings observed by Nakamoto (1952) may be more natural for the best π-orbital overlap. The 4-keto-pipecolic acid has its 4-keto group at a very favorable position to hydrogen bond with the phenolic hydroxyl group of the 3-hydroxy-picolinic acid. Conversely the 4-keto group may exist in the enolic form and hydrogen bond with the back side of the phenolic hydroxyl (which can, of course, be bonded with the neighboring —C=O of its own ring system). The proline ring and the two-carbon side chain of the D-α-amino butyric acid complete the side chain picture, with the latter fitting exactly in the suggested location. The N-methyl group of the substituted phenylalanine and the methyl group of threonine are also visible.

While viewing this "back" side of the staphylomycin model (Plate II (b)), it may be pointed out that ostreogrycin B (Eastwood, Snell & Todd, 1960) differs from staphylomycin in having a dimethylamino group attached to the benzene ring of N-methyl-L-phenylalanine at position Ⓐ. This additional substituent is very close to the nitrogen of the pyridine ring, and the two nitrogens would be close enough to transfer a charge if one of them were protonated. Steric hindrance by the two N-methyl substituents will direct the free electron pair of the dimethylamino group toward the pyridine nitrogen in this proposed ostreogrycin B conformation.

3. Arginine Vasopressin

The primary structure of arginine vasopressin has been adequately substantiated through the total synthesis of this compound (du Vigneaud, Gish, Katsoyannis & Hess, 1958). Although the compound contains only 8 amino acids, the bifunctional nature of the cystine and the C-terminal glycine amide supply a total of 9 amide bonds.

A proposed conformation for this hormone is presented in Plates III (a) and (b). The "front" or hydrophilic side (Plate III (a)) shows the hexagon on the right side of the photograph stabilized by the disulfide bond of the cystine unit. The stability of the partially completed hexagon on the left

side of the model may reside in hydrogen bonding forces between the guanidine portion of the arginine unit and one of the cystine sulfur atoms (see Plate III (b)) plus an additional bonding of the terminal glycine amide with one of the guanidine nitrogen atoms which will be discussed later. The latter hydrogen bond is more clearly seen in the open center of the left portion of the model in Plate III (a). This "front" side also shows the amino group of the N-terminal cystine in a central position of the right hexagon and allows a well-positioned hydrogen bond with the carbonyl oxygen of the asparagine amide group projecting into the open center of the hexagon from the "back" side. The cystine–asparagine amide junction is shown in the photograph (Plate III (a)) as a *cis* amide linkage. This particular amide arrangement together with the neighboring cystine carbonyl pointing into the center of the model would correspond to Fig. 2(a) and an oxygen–oxygen distance of about 5·2 Å. If the cystine–asparagine amide link is present as a *trans* amide configuration (—NH pointing inward, C=O pointing outward), then the carbonyl oxygen–oxygen separation corresponds to about 4·9 Å or slightly less (Fig. 1(b)). This *trans* arrangement would probably fit best in an aqueous environment at optimal biological conditions, and such an alternate model can be constructed without difficulty.

The "back" side of arginine vasopressin (Plate III (b)) with its several reactive side chains suggests a number of possible hydrogen bondings. In joining the disulfide linkage of the cystine unit, the two sulfur atoms have been placed in the most stable conformation with a dihedral angle of 90° between the two methylene groups (Calvin, 1954). The sulfur on the right hand side has one of its free electron pairs bonded to one of the hydrogen atoms of the asparagine amide group. The other sulfur electron pair may be joined to the tyrosine ring, either by the hydrogen atom of the phenolic hydroxyl or perhaps by the ring hydrogen atoms ortho to the phenolic hydroxyl. Another possibility would be that this second electron pair is available for interaction with other cell components or substrates in view of the possible disulfide interchange involved in the reactivity of this hormone (Schwartz, Rasmussen, Schoessler, Silver & Fong, 1960). The second hydrogen atom of the asparagine amide is bonded in this arrangement with the carbonyl oxygen atom of the glutamine amide. The carbonyl group of the asparagine itself is not clearly visible in either Plate III (a) or (b). Viewed from the perspective offered by Plate III (b), this —C=O unit points downward just below the asparagine amide's $—NH_2$ group; and this carbonyl oxygen, as already indicated, is in a position to hydrogen bond with the free amino group of the N-terminal cystine unit. The asparagine amide unit in this conformation, therefore, serves to bond two side-chain groups and the N-terminal amino group together. From analogue

studies of oxytocin, the glutaminyl[5]-oxytocin (Boissonnas, Guttmann, Jaquenod & Waller, 1956) was practically devoid of activity in six different biological systems (Berde, Doepfner & Konzett, 1957). Similarly, the conversion of the β-amide group of asparagine to the nitrile in arginine vasopressin itself (Katsoyannis, 1961) leads to an almost completely inactive compound; so that variations at this point may alter the activity considerably.

The sulfur atom on the left side in Plate III (b) has its two free electron pairs so orientated that they seem to be exactly positioned for bonding with the $—NH—C—NH_2$ portion of the planar guanidino group to form a six-membered hydrogen-bonded ring. In a similar manner the remaining imino group of the guanidine segment is hydrogen bonded with the back side electron pair of the nitrogen atom of the glycine amide, so that there may be two interconnected six-member bonded rings present in this portion of the model (see Fig. 4). The resonance stability of such an arrangement would be considerable.

One other possible steric effect may be mentioned. The suggested position of the paraffinic portion of the arginine side chain overlays the α-carbon atom of the glycine unit with its single hydrogen atom on the back side. This small unit offers no spatial conflict to the arginine side chain, whereas even a methyl group (e.g. C-terminal L-alanine amide) might cause steric hindrance at this point. A similar probable utility for minimal-sized sarcosine units was suggested in our proposed actinomycin conformation (Warner, 1961) and will also be mentioned in our discussion of the Try-Gly junction in Plate IV (b).

4. ACTH Decapeptide

The N-terminal decapeptide of the ACTH molecule contains the amino acid sequence:

1 2 3 4 5 6 7 8 9 10
Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Try-Gly

The last seven components of this sequence are also common to the melanotropins and β-MSH (Li, 1957). In constructing the conformation pictured in Plates IV (a) and IV (b), the assembly of the model was begun with the N-terminal serine (Ser-1). The terminal amino group was placed so that it pointed inward towards the center of the hexagon. In that position it is able to form a hydrogen bond with one of the electron pairs of the methionine sulfur (Met-4). The main features of the "front side" of the model (Plate IV (a)) are the N-terminal amino group in the center of the right hexagon and the guanidino group of arginine in the left hexagon. The Glu-His amide junction is indicated in the *trans* form to permit the carbonyl oxygen–oxygen distance of Fig. 1(b). (A similar consideration

was pointed out for the cystine–asparagine amide in Plate III (a); and I presume that if the amide bond has a certain degree of "double bond" character, it will be of little consequence whether the double bond has its initial origin in a *cis* or *trans* configuration.) In a model constructed from atoms having all of the exact bonding angles, perhaps the hydrogen of this amide might be in a position to bond with the N-terminal amino group in the center of the hexagon. In Plate IV (a) the carboxy-terminal Gly-10 has been joined to an amide nitrogen which is really a part of the Lys-11 unit. This —NH— group is shown in a hydrogen bond with one of the nitrogen atoms of the guanidino grouping of Arg-8. This type of bonding of hydrogen to nitrogen could perhaps account for some of the twenty hydrogen atoms in ribonuclease which are not readily displaced by deuterium (Scheraga & Hermans, 1960) although Donohue (1952) indicates that the —NH—N— bonding is relatively weak.

The "back side" of the ACTH decapeptide (Plate IV (b)) shows the large number of possible side-chain interactions which this conformation permits. First of all, the hydroxyl group of Ser-1 is in a position for very close interaction with the γ-carboxyl group of Glu-5, even allowing ester formation, which would be a possible reaction in the hydrophobic surroundings. This potential ester bond has actually been joined in the model to aid in the assembly. This linkage holds the γ-carbonyl (or carboxyl) oxygen in such a position that one of its electron pairs may be bonded with the phenolic hydroxyl of the Tyr-2 unit (Edsall & Wyman, 1958), and the direction of the hydroxyl's approach to the carbonyl oxygen is very favorable. Proceeding along the peptide chain, the hydroxyl group of Ser-3 is bonded to one of the electron pairs of the Met-4 sulfur atom. (The other electron pair of this sulfur atom is bonded, as previously indicated, to the N-terminal amino group of Ser-1.) The γ-carboxyl group of Glu-5 is not only joined to the hydroxyl group of Ser-1 and hydrogen-bonded through its oxygen to the phenolic hydroxyl of Tyr-3, but its carbonyl *carbon* atom is touching one of the nitrogen atoms of the imidazole ring of His-6. In this position it would be capable of existing as a labile grouping of the acylimidazole type. The possible interrelations between imidazole, serine hydroxyl, and carbonyl oxygen which have been suggested by Cunningham (1957), Bender, Schonbaum, and Hamilton (1961), and Krupka and Laidler (1961) are thus clearly compatible with this portion of the proposed conformation. The imidazole ring is held in position not only by this potential contact with the side chain of Glu-5, but also by the neighboring planar benzene ring of Phe-7. The side-chain of the Arg-8 lies between the same benzene ring of Phe-7 and the indole nucleus of Try-9. In the suggested conformation, the imino nitrogen of the guanidine unit may serve as a bonding partner between the indole ring —NH— and

one of the electron pairs of the Tyr-2 phenolic oxygen moiety, although the exact arrangement of the three nitrogen atoms of the guanidine unit cannot be rigidly decided from the models. It will be observed that all of the bonding atoms in this arrangement can approach the members to which they bond at almost exactly the best angle for the most stable hydrogen bond. It is also possible that the other amino nitrogen atom of the Arg-8 guanidine can participate in the sort of serine–carboxyl interaction proposed by Erlanger (1960). Plate IV (a) shows this —NH_2 group of the guanidine moiety pointing directly toward the junction of the Ser-1 hydroxyl and the γ-carbonyl of Glu-5. Both the γ-carbonyl oxygen (below the visible surface) and the serine hydroxyl (just below the His-6 carbonyl group) have the necessary additional electron pairs to bond the —NH_2 group in the manner proposed by Erlanger. The position of Gly-10 (scarcely visible in Plate IV (b)) in the amino acid sequence at this point may be another example of the use of this minimal-sized unit to avoid steric interference. The Try-9 side chain passes directly over the Gly-10 α-carbon atom with its single α-hydrogen on this back side. This single hydrogen atom fits into the space available in the indole moiety provided by the absence of a hydrogen atom at the 4-position of the indole ring.

5. Concluding Remarks

In our previous paper (Warner, 1961) three polypeptides having a large preponderance of hydrophobic side chains were discussed, and it was suggested that the stability of these conformations could reside in the hydrophobic interactions of the side chains and the probable water bonding between adjacent carbonyl oxygens of the peptide chain. In this present paper, the discussion of the water system has been extended to show that its "second neighbor" oxygen distribution has a planar hexagonal pattern which is very similar in shape and dimension to the carbonyl oxygen pattern of our suggested hexagonal peptide units. The hexagonal concept has now been studied in polypeptides having more reactive side chains, and models of the proposed conformations have shown that numerous bonded interactions are capable of existing, some of them in accord with suspected or established side-chain reactivities. It will thus be apparent that the suggested tertiary structures of these known polypeptides can provide a large number of stabilizing factors, especially in the aqueous system where the water pattern itself may be a strong force by virtue of the postulated orderly interaction with the peptide bonds. However, if a proposed conformation is to be regarded as a favorable one, this high degree of stability must be coupled with a high degree of potential reactivity within the molecular arrangement. I would like to suggest that the conformations proposed for the several polypeptides are also consistent

with molecular arrangements of high reactivity. To illustrate these possibilities for reactivity, the reader is referred to the discussions on charge transfer by Szent-Györgyi (1960) in which he indicates that reactivity may be associated with structures having a good capacity for electron or charge transfer. In this same discussion he indicates that electron transfer may take place best in a hydrophobic medium from which water molecules are excluded. I previously mentioned this point briefly in our proposed conformation of actinomycin (Warner, 1961) where the resonating chromophore is surrounded by hydrophobic, electrically insulating groups which could confine the electron transfer to the definite pathway of the resonating system. A brief discussion of electron flow in biological systems is also given by Dash (1961).

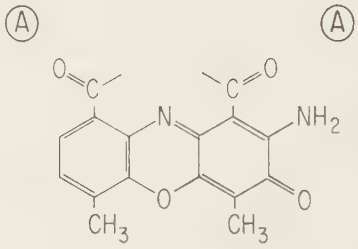


Fig. 3. Electron transfer pathway in the actinomycins.

Returning to the concept of single or multiple hexagonal units in the reactive polypeptides, I have proposed (Warner, 1961) that the amide or hydrophilic "front side" of the models may constitute the reactive surface. Within each hexagonal unit, a reactive group was shown to be present at the center that could contact substrate molecules in the aqueous phase. The contact between the center group and the substrate molecule would, of course, alter the reactive center (e.g. the amino group would be protonated by the approach of a carboxyl group). If such a reactive center is in contact with an electron transfer pathway formed by side-chain interactions, then the charge may be transferred by this medium to another hexagonal center where it is capable of producing an effect on another part of the substrate molecule. Since the actual charge transfer occurs within the confines of the "back side" hydrophobic regions of the reactive polypeptide, it may reappear in the hydrophilic "front side" only at certain fixed positions which could be generally related to the reactive centers of the various hexagonal units. The electron transfer pathways between two such reactive centers are quite clear in the actinomycins (Fig. 3) and arginine vasopressin (Fig. 4), where the Ⓐ symbols indicate the groupings which appear in the hexagonal centers of the hydrophilic surface of the peptide molecule.

With regard to the charge transfer pathways between the two γ-amino ornithine groups in gramicidin S (Warner, 1961), the possibility that the two benzene rings properly positioned between the amino groups could

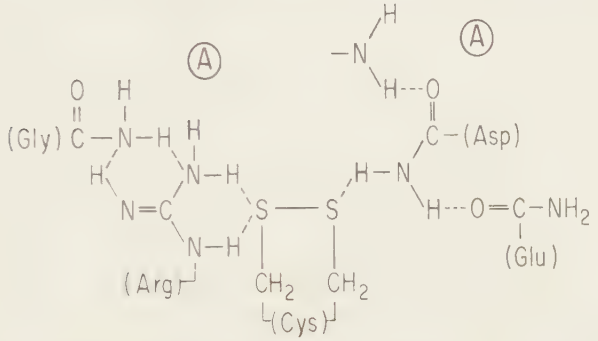


FIG. 4. Electron transfer pathway in arginine vasopressin.

serve in this capacity is perhaps more attractive since the disclosure of the conductivity of polymers of cyclopentadiene in nonpolar solvents from the work of Wasserman (1960).

The conformations for etamycin (Warner, 1961), staphylomycin, and ostreogrycin B (Plates II (a) and II (b)) also have quite elaborate charge

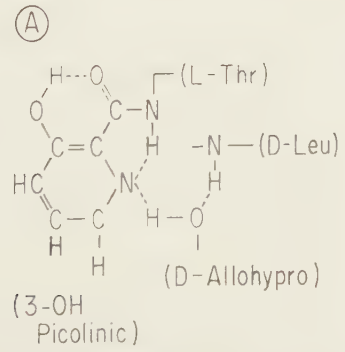


FIG. 5. Electron transfer pathway in etamycin.

transfer pathways through the suggested side-chain interactions, although in these instances, the activity seems to revolve around a single active center. The probable transfer linkages are indicated for two of these models in Figs. 5 and 6. These charge transfer pathways may also serve to transfer electron effects to underlying layers of the cell structure.

Charge transfer pathways are also present in the ACTH decapeptide model, but it is difficult to present them in their proper relation in a two-

dimensional diagram. Some of them have already been pointed out in the discussion of the stability factors in the ACTH sequence (Plates IV (a) and IV (b)); for the hydrogen bondings which have been described mostly

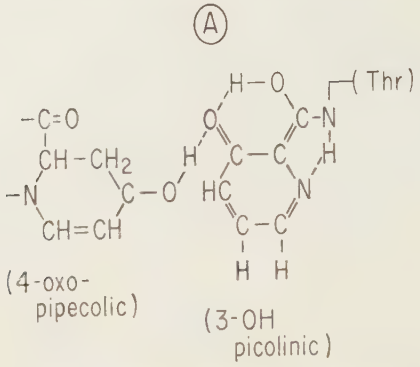


FIG. 6. Electron transfer pathway in staphylomycin.

in terms of their stabilizing influence are, of course, simultaneously an important aspect of charge transfer mechanistics. Even the —NH-group of the indole nucleus, by its close proximity to and hydrophobic fitting with the arginine side chain, could conceivably participate in a "one electron" transfer system, although this substance is not usually regarded as an oxidation–reduction agent in the classical sense. A further problem in attempting to trace the electron pathways in a partial structure such as the ACTH decapeptide resides in the fact that some incomplete transfer pathways may eventually terminate in probable hexagonal units formed by later amino acid units in the chain. Consequently, a more profitable examination of charge transfer in ACTH can be conducted when we are able to obtain enough atom models to complete the entire sequence following our suggested hexagonal plan. This work is now in progress, as well as similar studies with TMV protein.

The whole concept of charge transfer involving fixed electron pathways within the hydrophobic regions of the peptide or protein molecule is a very attractive one from the standpoint of a possible insight into the mechanism of action of peptide structures. By such pathways one portion of the substrate can initiate the electron transfer which produces reaction with another portion of the substrate, thus furnishing the energy for its own hydrolysis, synthesis, or oxidation.

Similar conformations have been constructed with the N-terminal portions of other biologically active peptides and additional antibiotics, but the present examples are sufficient to indicate the structural and functional possibilities of the proposed hexagonal conformations.

Although the main emphasis of this paper has suggested a maximum stabilizing effect of water on peptide or protein structures at about 37°, biological processes occur to varying extents over a range near this temperature. It may be significant that the carbonyl oxygen–oxygen distances represented by Fig. 1(a) (4·76 Å) and Fig. 1(b) (4·90 Å) correspond to "second neighbor" oxygen distances for water at about 25° and 65° respectively. The results of Scheraga (1960) and von Hippel & Harrington (1960) present the possibility of some type of conformational stability for the proteins studied within the 30°–55° range, as evidenced by plateaus in the specific rotation values. It is tempting to speculate that the 37° temperature may stabilize the peptide bond in the —C(OH)=N— form intermediate between the *cis* and *trans* positions; but that considerable conformational stability could conceivably be conferred by the aqueous system over the entire temperature range allowed by the *cis–trans* oxygen interval. The interesting studies on ionic dissociation at the peptide linkage (Browne, 1961) and the entire problem of the pH effect on transition temperatures may also be pertinent to the discussion.

I would like to acknowledge the advice and encouragement of Professor C. H. Li in the development of this work. I would also like to thank members of the Upjohn staff for many helpful conversations and suggestions, particularly Drs. S. H. Eppstein, E. L. Schumann, D. T. Gish, A. J. Parcells, and H. Ko. The plates and figures were prepared with the cooperation of Mr. E. E. Beals of the Research Photography group.

Mr. J. E. S. Whitney of Catalin Limited kindly supplied the special peptide nitrogen atoms for use with their regular model sets in assembling these conformations.

REFERENCES

Bender, M. L., Schonbaum, G. R. & Hamilton, G. A. (1961). *J. Polym. Sci.* **49,** 75.
Berde, B., Doepfner, W. & Konzett, H. (1957). *Brit. J. Pharmacol.* **12,** 209.
Boissonnas, R. A., Guttmann, St., Jaquenod, P. A., Waller, J. P. (1956). *Helv. chim. acta.* **39,** 1421.
Browne, P. F., (1961). *Nature, Lond.*, **189,** 658.
Calvin, M. (1954). *Fed. Proc.* **13,** 703.
Corey, R. B. & Pauling, L. (1953). *Proc. roy. Soc.* **B141,** 10.
Cunningham, L. W. (1957). *Science,* **125,** 1145.
Dash, H. H. (1961). *Science,* **133,** 1653.
Donohue, J. (1952). *J. phys. Chem.* **56,** 502.
du Vigneaud, V., Gish, D. T., Katsoyannis, P. G. & Hess, G. P. (1958). *J. Amer. chem. Soc.* **80,** 3355.
Eastwood, F. W., Snell, B. K. & Todd, A. R. (1960). *J. chem. Soc.*, 2286.
Edsall, J. T. & Wyman, J. (1958). "Biophysical Chemistry", Vol. I, p. 62. Academic Press, New York.
Erlanger, B. J. (1960). *Proc. nat. Acad. Sci., Wash.* **46,** 1430.
Katsoyannis, P. G. (1961). *J. Polym. Sci.*, **49,** 64.
Kauzmann, W. (1959). "Advances in Protein Chemistry", **14,** p. 33 ff. Academic Press, New York.
Krupka, R. M. & Laidler, K. J. (1961). *J. Amer. chem. Soc.* **83,** 1445.

Lange, N. A. (1956). "Handbook of Chemistry", 9th Edition. Handbook Publishers Co., Sandusky, Ohio.
Li, C. H. (1957). "Advances in Protein Chemistry", **12**, 269. Academic Press, New York.
Morgan, J. & Warren, B. E. (1938). *J. chem. Phys.* **6**, 666.
Nakamoto, K. (1952). *J. Amer. chem. Soc.* **74**, 1739.
Scheraga, H. (1960). *Brookhaven Symposia in Biology, Protein Structure and Function*, **13**, 73.
Scheraga, H. A. & Hermans, J. J. (1960). *J. Amer. chem. Soc.* **82**, 5156.
Schmidt, G. M. J., Hodgkin, D. C. & Oughton, B. M. (1957). *Biochem. J.* **65**, 744, 752.
Schwartz, I. L., Rasmussen, H., Schoessler, M. A., Silver, L. & Fong, C. T. O. (1960). *Proc. nat. Acad. Sci., Wash.* **46**, 1288.
Sheehan, J. C., Zachau, H. G. & Lawson, W. B. (1957). *J. Amer. chem. Soc.* **79**, 3933.
Szent-Györgyi, A. (1960). "Introduction to a Submolecular Biology". Academic Press, New York.
Vanderhaeghe, H. & Parmentier, G. (1959). *Bull. Soc. chim. Belg.* **68**, 716.
Vanderhaeghe, H. & Parmentier, G. (1960). *J. Amer. chem. Soc.* **82**, 4414.
Von Hippel, P. H. & Harrington, W. F. (1960). *Brookhaven Symposia in Biology, Protein Structure and Function*, **13**, 222.
Warner, D. T. (1961). *Nature, Lond.* **190**, 120.
Wasserman, A. (1960). *Chem. Engng. News*, **38**, (#51), 46, Dec. 19.

J. Theoret Biol. (1961) **1**, 529–534

Ion-Exchange Mechanisms in the Membranous Labyrinth: a Suggested Basis for the Sudden Attacks in Ménière's Disease

L. Naftalin and M. Spencer Harrison

County Hospital, Lincoln, England

(*Received* 28 *March and in revised form* 19 *July*, 1961)

Although the triad of symptoms of vertigo, tinnitus and deafness was a fairly common affliction and had been known for a century before his time, it was not until 1861 that Ménière published his well-known description. The case was that of a young girl who, following exposure to cold, was attacked by sudden deafness, intense vertigo, vomiting and pyrexia. Post-mortem examination revealed a reddish-yellow coagulum within the labyrinth, and it has since been customary to regard the case as an instance of acute haemorrhagic labyrinthitis. His nine subsequent cases were probably not of infectious origin (Simonton, 1940).

Since that time Ménière's name has come to be attached somewhat loosely to a variety of clinical conditions presenting the syndrome of vertigo, tinnitus and deafness. In some of these, signs and symptoms of cochlear and vestibular upset are clearly recognizable as being due to inflammatory, neoplastic or other processes involving the labyrinth, the eighth nerve, or its central connections. With these excluded, however, there still remains a considerable group in which these signs and symptoms present certain recognizable peculiarities in their mode of onset, character and clinical course. The papers of Crowe (1938), and of Wright (1937) did much to clarify our views on the clinical identification of this group of subjects. Both of these authors were agreed in attributing the condition to a specific variety of labyrinthine disease, and their views were strongly supported by the publication in 1938 and 1940 by Hallpike, Cairns and Wright of the histological findings in the temporal bones of three cases of this kind. In all three, the affected labyrinth was found to be the seat of certain peculiar changes indicative of distention of the endolymph system. These findings have since been fully confirmed and it is now possible to regard these investigations as having established the morphological basis of a disease *sui generis* of the labyrinth. According to modern views based upon these clinical and pathological studies, the term "Ménière's disease"

should be reserved for this particular group of cases readily recognizable as a clinical entity and presenting within the labyrinth the specific type of morbid change indicative of endolymphatic distention.

The clinical observation that a low salt intake was of value in reducing the frequency and intensity of the attacks of vertigo (Dederding, 1929; Mygind & Dederding, 1938; Furstenberg, Lashmet & Lathrop, 1934; Furstenberg, 1936) preceded the histological demonstration of hydrops of the endolymphatic spaces (Hallpike & Cairns, 1938). It was concluded from these combined observations that a sodium retention was occurring as a result of a sodium intake considered to be excessive for this type of patient, though this intake would not necessarily be greatly in excess, if at all, of a normal intake. In modern terms these clinical observations could be interpreted to mean that the maintenance of a high level of aldosterone secretion was required for the suppression of the acute symptomatology.

Perlman, Goldinger & Cales (1953) and Harrison & Naftalin (1958) noted that the real retention of sodium induced by the use of DOCA did not increase the frequency or intensity of the acute episodes of the disease: indeed in both series of cases a number of patients claimed to be greatly improved. On reviewing the data we realized that DOCA was a long-acting sodium-retaining steroid which interfered with the natural cycle of sodium homeostasis. We therefore repeated and extended our observations on the next series of patients using aldosterone to induce short periods of sodium retention (Naftalin & Harrison, 1961).

Our findings have been that no attack has been precipitated by injections of aldosterone with or without a sodium load in patients suffering from Ménière's disease, just as with injections of DOCA. Nevertheless, a number of objectively observed attacks have been noted, not while aldosterone was being administered but while a load of sodium chloride was being given. Data were gathered so that the trend of the cumulative balance for sodium could be (subsequently) traced out in order to determine in which phase of the continuing homeostatic cycle the attack actually occurred. It was found that the phase was that of the downstroke, or sodium diuresis, although the total body sodium at that moment could not be assessed with respect to the individual's average "normal" total body sodium. This repeated observation of a sodium diuresis, despite a continuing sodium chloride loading, suggests that a decrease in aldosterone secretion was occurring, possibly as a result of the sodium loading. This would, indeed, be no more than an exaggeration of the normal homeostatic cycle. Our new conclusion, therefore, was that attacks of vertigo coincided with a decrease in aldosterone secretion, and this correlates well with the older clinical observations that a suppression of symptoms occurred most often

with a "salt-free" diet, i.e. with maintenance of a high level of aldosterone secretion.

In an earlier paper (Naftalin & Harrison, 1958) we were led, by considering the results of the electron microscope studies of Pease (1956), Palade (1956) and Smith (1957), to the suggestion that the stria vascularis is a membrane having some very similar characteristics to those of the renal tubular cells. Aldosterone is known to exert its action on kidney, tubular, intestinal and salivary gland cells by facilitating reabsorption of Na^+ at these sites, usually with K^- excretion. Since the endolymph is a fluid having a high K^- concentration, it appeared a reasonable assumption that the stria vascularis is the site of a Na^-/K^- exchange pump. This assumption enabled a dynamic circulation of labyrinthine fluids to be described.

The data assembled in the experiments yielding information on the phase of the sodium homeostatic cycle also included daily body weights, the observations being made under closely controlled conditions. The trend shown by the weight curves suggests that at the time of a Ménière's attack, body weight was either stationary or actually increasing, i.e. in some cases water retention was in progress. This "out of step" relation of the sodium and water homeostatic mechanisms can also be seen in data from a normal subject, but is quickly corrected and, in any case, such changes may be presumed normally not to be in excess of the osmotic tolerance of the tissues. It seems possible that this may not be true of the site of the lesion in Ménière's disease.

The lesion in the end-organ is conceived as being, in the first instance, small in extent, leaving the tissue involved with the greater part of its normal "reserve" for tolerating any single minor assault. This postulate derives from the clinical picture of the early attacks which are acute and from which usually complete and rapid recovery is made. The normal subject tolerates a small "out-of-step" relation of water and salt homeostasis: it is common experience also to have a transient labyrinthine disturbance as a result of sudden change of posture. The long periods of remission in Ménière's disease would seem to justify an approach which assumes that the labyrinthine tissue involved can, in these cases, tolerate "normal" degrees of homeostatic imbalance, or of postural, vasomotor or other labyrinthine stimulus. When, however, these stimuli are coincident and additive or somewhat excessive an attack of vertigo is a likely outcome.

On the basis of this hypothesis we have developed a regime which attempts to smooth out the fluctuations of the homeostatic cycle, thus presenting the labyrinthine ion exchange structures with a minimum of disturbance or of work to do. The regime prescribes a small NaCl addition (particularly if the patient has a naturally low salt intake) and a larger

addition of KCl. The preponderance of K^+ is intended to maintain an (increased) output of aldosterone, while the smaller amount of Na^+ should ensure that any dietary fluctuation of Na^+ intake is reduced to a ripple. This is an important point in treatment since this regime is easy to maintain compared with the alternative diet of strict low salt intake, a rather miserable procedure for the patient.

It follows from the above that the regime described may prevent attacks, maintain a steady state or even induce remissions when a cycle of attacks has started (and our experience of the past eighteen months supports this contention), but cannot, of itself, treat the fundamental lesion of Ménière's disease.

The fundamental lesion of Ménière's disease, affected by the water and electrolyte changes, lies in the end-organs in the inner ear. This has been determined by clinical observation and surgical experience. The end-organs are bathed in, or are intimate contact with, the endolymph with its high K^+ concentration. The following discussion attempts to connect the clinical and laboratory findings described above with the physiology and biochemistry of the inner ear. This latter subject was reviewed by Davis (1957) who drew attention to the circumstance that no satisfactory explanation had yet been found for the high K^+ concentration in the endolymph. There appears to be no correlation of this high potassium concentration with the endolymphatic DC potential since the potential in the cochlear duct is + 67 mV compared with + 4 mV in the utricle and + 1 mV in the saccule (Smith, David, Deatherage & Gessert, 1958).

Further work on the potentials in the labyrinth has recently been reported by Trincker (1959) who found that ampullary endolymph, in the guinea-pig was + 40 mV measured against the perilymph, but at the surface of the cupula a higher positive potential, 70 mV, existed. Potentials in the deeper layers of the cupula varied with "pockets" of weak negative (− 6·5 mV) potentials separated spatially by distances equal to those between the hair processes.

There is evidence that the tectorial and otolithic membranes and the cupulae are of the nature of a glycoprotein, probably sulphur-containing, although mucoprotein is not ruled out as a possibility (Wislocki & Ladman, 1955). In a section of the cochlear duct the tectorial membrane is seen to overlie the organ of Corti, separated from the bodies of the hair cells by the reticular lamina. The tunnel of Corti and space of Nuel are filled with perilymph (Davis, 1957) but the internal sulcus is part of the endolymphatic system. The tectorial membrane is, therefore, maintained in a high potassium medium (as are also the cupulae) but, as it is material lacking cellular structure, it cannot effect a metabolic exchange of sodium and potassium and acts only after the fashion of a resin. The continuous

recharging with potassium would have to be effected elsewhere in the endolymphatic system.

Engström and Wersäll (1958) conclude from their histological and electron microscope studies that the hair processes of the hair cells protrude through the reticular lamina into the tectorial membrane. It is generally accepted that a major transformation of the mechanical energy of the sound waves in the fluid systems of the inner ear into chemical energy leading to the electrical energy of the action potential in the nerve cell, occurs in or around the hair cell and hair processes. The body of the hair cell is bathed in the usual extracellular fluid, but the hair process (the presumed sensitive element) lies in a high potassium medium. Unless the cuticle of the hair process is completely impermeable to sodium and potassium the membrane potential in the hair process must be different from that of the body bathed in extracellular fluid.

There are several ways in which the physical wave-energy in the endolymph might be transduced to the electrical energy of the nerve cell, but whichever mode is considered most likely, the tectorial membrane and the cupulae by virtue of their structure in respect both of their physical properties and of the ionic environment they provide for the hair processes, are important as mediators in the response of the hair cells to mechanical stimulus.

The physical nature of these materials is that of a jelly, stiffened by a fibrillar structure and giving the histological staining reactions of a glyco- or mucoprotein. Such a molecule would have the properties of a polyelectrolyte and would exhibit characteristics described by Orofino and Flory (1959). The property of greatest interest in connection with the tectorial membrane and cupulae is that, with a polyelectrolyte chain, the osmotic effects of small ions (e.g. potassium or sodium) are greatly depressed by the charges on the chain, particularly for high degrees of neutralization, a condition which normally obtains in the fluids of the inner ear.

If, in Ménière's disease, the sudden stimulation of the balancing mechanism is mediated through an osmotic change, a metabolic breakdown from anoxia of the sensitive cells would neither be fast enough nor quantitatively large enough to produce a sufficient number of particles, and moreover, rapid recovery of the sensitive cells would hardly be possible. The nature of the attacks of dizziness in Ménière's disease suggest that the changes in the endolymph are either acute, reversible pressure changes, or acute, reversible electrical, i.e. ion transport, changes. Either or both of these physical phenomena would occur instantaneously with a critical change in the degree of neutralization of the polyelectrolyte (tectorial membrane or cupulae). This critical change in the degree of

neutralization of the polyelectrolyte could come about by a mild degree of local acidosis, as a result of local, temporary anoxia, produced by a vasospasm induced humorally or by postural reflex. Beyond the critical pH, H^+ would displace Na^+ or K^+, the Na^+ being the more readily dissociated. The released cation accompanied by an anion (e.g. HCO_3^- from the hair cells) would sharply affect osmotic pressure, and electrical equilibria within the polyelectrolyte "environment" would also be disturbed. It is suggested, therefore, that the high K^+ concentration in the endolymph is related to the probable polyelectrolyte character of the tectorial membrane and cupulae, and is involved in the electrical structure and function of these components of the transducing system in a way that the larger hydrated, and more easily displaced, Na^+ could not so satisfactorily fulfil.

We wish to thank Dr. A. K. Pittman, Director of Clinical Research, CIBA Laboratories Ltd., for the supply of aldosterone, and for financial assistance for technical expenses.

REFERENCES

Crowe, S. J. (1938). *Medicine, Baltimore*, **17**, 1.
Davis, H. (1957). *Physiol. Rev.* **37**, 1.
Dederding, D. (1929), *Acta Otolaryng., Stockh. Suppl.* x-xi.
Engström, H. & Wersäll, J. (1958). *Int. Rev. Cytol.* **7**, 535.
Furstenburg, A. C. (1936). *Trans. Pacif. Cst. oto-ophthal. Soc.* **21**, 150.
Furstenberg, A. C., Lashmet, F. H. & Lathrop, F. (1934). *Ann. Otol. etc., St. Louis*, **43**, 1035.
Hallpike, C. S. & Cairns, H. (1938). *J. Laryng.* **53**, 625.
Hallpike, C. S., Cairns, H. & Wright, A. J. (1940). *J. Laryng.* **55**, 59.
Harrison, M. S. & Naftalin, L. (1958). *Proc. R. Soc. Med.* **51**, 624.
Mygind, S. H. & Dederding, D. (1938). *Ann. Otol. etc., St. Louis*, **47**, 55, 768.
Naftalin, L. & Harrison, M. S. (1958). *J. Laryng.* **72**, 118.
Naftalin, L. & Harrison, M. S. (1961). *Lancet*, ii, 106.
Orofino, T. A. & Flory, P. H. (1959). *J. phys. Chem.* **63**, 283.
Palade, G. E. (1956). *J. biophys. biochem. Cytol.* **2**, (Suppl.) 85.
Pease, D. C. (1956). *J. biophys. biochem. Cytol.* **2**, (Suppl.) 203.
Perlman, H. B., Goldinger, J. M. & Cales, J. O. (1953). *Laryngoscope, St. Louis*, **63**, 640.
Simonton, K. M. (1940). *Ann. Otol. etc., St. Louis*, **49**, 80.
Smith, C. A. (1957). *Ann. Otol. etc., St. Louis*, **66**, 521.
Smith, C. A., David, H., Deatherage, B. H. & Gessert, C. F. (1958). *Amer. J. Physiol.* **193**, 203.
Trincker, D. (1959). *Ann. Otol. etc., St. Louis*, **68**, 145.
Wislocki, G. B. & Ladman, A. J. (1955). *J. Anat.* **89**, 3.
Wright, A. J. (1937). *Proc. R. Soc. Med.* **31**, 87.

J. Theoret. Biol. (1961) 1, 535–541

The Biological Designing of Antimetabolites for Malignant Growth

G. F. Gause

Institute of Antibiotics, Academy of Medical Sciences, Moscow

(*Received 7 May* 1961)

In micro-organisms mutants have been produced with biochemical alterations that are similar to but not identical with the corresponding biochemical alterations in the cells of malignant tumours of higher organisms. As is known, metabolites of some biochemical systems show antimetabolite properties when tested in similar biochemical systems of other species of organisms. In view of this principle of antimetabolite inhibition, biochemical components of micro-organisms "analogous" to cancer cells should be studied in greater detail as possible antimetabolites in the similar but not identical biochemical systems of cancer cells.

Introduction

The theoretical approach to designing and improvement of anti-tumour drugs is being developed at present in two main directions. Firstly, when the active chemical grouping possessing anti-tumour action is already known (e.g. nitrogen mustard), the selectivity of its action upon various types of tumours can be considerably enhanced by building it into a more complex molecule in a manner suggested by various biochemical ideas. Toxic nitrogen mustard can be converted into a less toxic derivative, from which the original toxic compound could be liberated only by enzymic action directly in the cells of a tumour (Danielli, 1959). Combining nitrogen mustards with peptides, one can enhance the selectivity of their action upon specific types of tumours (Larionov, 1959; Knuniants, Golubeva & Kildisheva, 1960).

Secondly, the search for new active chemical groupings possessing anti-tumour action can be based on the principle of antimetabolites; by synthesis one can prepare chemical analogues of essential metabolites of tumour and normal cells, which would interfere with function of metabolites and in some cases could inhibit malignant growth (Woolley, 1954). The main difficulty for work along these lines is to attain a greater selectivity of action of synthetic antimetabolites for the cancer than for the

normal cells of the organism. This is directly related to the fact that no specific metabolites operating only in tumour cells are known at present.

It is important to note that only a small number of metabolites is known in contrast to the actually functioning number which may be very large. In view of the insufficient development of biochemistry, the important idea of making antimetabolites by chemical synthesis cannot be realized now to the full degree.

In this paper a new possible approach to obtaining antimetabolites of tumour cells will be discussed, with the idea of attaining this goal not by chemical synthesis but by means of producing biochemical analogues of tumour cells in micro-organisms and isolating from them antimetabolites that inhibit the growth of tumour cells. This new line of investigation is based upon two concepts: (1) the possibility of producing in micro-organisms some mutants with biochemical alterations in the cells which are similar to but not identical with corresponding biochemical alterations in the cells of malignant tumours of higher organisms; (2) the fact that progression from metabolite action to antimetabolite potency can be attained with small change in the structure of a molecule, and the fact that metabolites of some biochemical systems show antimetabolite properties when tested in similar biochemical systems of other species of organisms. Figure 1 presents a scheme showing the biological approach to the designing of antimetabolites for malignant growth which is based upon the two concepts listed above. The elements of this scheme will be discussed in greater detail in the following paragraphs.

Biochemical Systems of Cancer Cells

The biochemical characteristics of tumour cells are still controversial and there are many gaps in our knowledge in spite of numerous investigations. The deficiency of respiration in tumours, discovered by Warburg about thirty years ago and recently reconfirmed (Warburg, 1930, 1956), still remains an important generalization, in spite of some criticisms (Weinhouse, 1956). Unfortunately, at present one cannot specify exactly what respiratory enzymes are defective, and how balances of these enzymes are distorted in tumour cells. Recent observations on deficiency of cytochromes b and c_1 in cells of some tumours are of interest (Monier, Zajdela, Chaix & Petit, 1959).

The deficiency of respiration represents only one aspect of malignant transformation of the cell, and is accompanied by deficiencies in a number of other enzymatic systems. Deletion in tumour cells of specific enzymes that degrade purines, pyrimidines and their derivatives has been recorded (Bennett, Skipper, Simpson, Wheeler & Wilcox, 1960), and among these the low content of xanthine oxidase has been studied in detail and dis-

cussed (Haddow, de Lamirande, Bergel, Bray & Gilbert, 1958). In hepatoma and other tumours a number of enzymes are missing or very low, notably catalase, arginase, cytochrome-c-reductase, tyrosine-transaminase, and others (Potter, 1958). In view of the variety of deficiencies in enzymatic systems of tumour cells, caution should be observed in assigning prime significance in malignant transformation to any one deficiency.

The deletion of enzymes in malignant cells is accompanied by the deletion of tissue-specific antigens (Weiler, 1959; Zilber, 1960). The biochemical nature and functional significance of these antigens is so far unknown. It has been mentioned that the loss of immunologically specific functions of tumour cells might be related to deletions of some tissue-specific surface proteins, and that the loss of these could lead to metastasis and invasion of tumour cells into the surrounding tissues (Potter, 1958). In the final analysis, the hereditary enzymatic and immunological defects of tumour cells might be related to deletions in molecules of deoxyribonucleic acid, as far as the latter are important determinants of hereditary characteristics of cells.

In complex cellular organelles containing aggregates of enzymes, the loss of some functions would lead to a transformation of the organelle which might be recognized as a gain or appearance of some new functions. This concerns, for instance, the observations of Zilber, Abelev, Avenirova, Engelhardt & Baidakova (1959), about the appearance in granules of mouse hepatoma of a specific antigen which is lacking in mouse liver cells. In the biochemical field these ideas have been developed by Woolley (1953). He recorded the evidence for the synthesis of vitamin B_{12} by spontaneous tumours of mice, and considered this as a special case of a more general biological principle that the undifferentiated tissue differs from the differentiated in possessing the ability to synthesize a particular biologically important compound.

Biochemical Systems of Analogues of Cancer Cells in Micro-organisms

Biochemical mutants with imparied respiration and a distorted system of cytochromes have been obtained in number of micro-organisms: yeasts *Saccharomyces* and *Candida*, fungus *Neurospora*, bacteria *Staphylococcus aureus*, *Bacterium coli*, *B. paracoli*, *Bacillus mycoides* and others, as well as in a flagellate, *Polytoma*. A review of this field has been published recently (Gause, 1959) and will not be repeated here.

Mutants with impaired respiration in staphylococci induced by ultraviolet radiation are of considerable interest, as far as their growth is selectively inhibited by a number of synthetic and natural compounds, which are used at present in the field of cancer chemotherapy. Such antibacterial antibiotics as tetracyclines, gramicidin S and novobiocin inhibit

the growth of parent staphylococci and their mutants with impaired respiration to the same extent, while in respect of penicillin, streptomycin, neomycin and albomycin mutants are somewhat more resistant as compared with parent cells. An entirely different picture is observed when a number of anti-cancer preparations is assayed upon these respiration-deficient mutants of staphylococci. It has been recorded that alkylating agents (chloroethylamino-phenylalanine, chloroethylamino-phenylamino-butyric acid, chloroethylcarbamoyl-serine, leukeran, degranol, myleran-mannitol, miracil D, triethylenemelamine), antimetabolites of nucleic acids (6-mercaptopurine, 6-chloropurine, 5-fluorouracil, 6-azauracil-riboside), and anti-cancer antibiotics (actinomycin C, mitomycin C, actinobolin, actidione) selectively inhibit the growth of mutant staphylococci with respiration deficiency (Gause, 1960; Gause & Kochetkova, 1960). In this connection the available information on biochemical systems and metabolites of these respiration-deficient mutants of staphylococci should be discussed in greater detail.

It was recorded that respiration coefficients in mutant staphylococci with impaired oxidation are decreased by 40 to 60% as compared with parent values, and respiration in mutants is less sensitive to cyanide inhibition. The cytochrome pattern is also altered in mutants. Normal staphylococci reveal three bands of cytochromes: a_2, a_1, b_1; in mutants the band a_2 is absent, and instead of band b_1 two new bands are observed. In respect of splitting of band b_1 of the cytochromes, the mutants of staphylococci with impaired respiration are reminiscent of the respiration-deficient mutants of *B. coli* (Gause, Ivanitskaia & Vladimirova, 1958).

Further studies have shown that various biochemical changes accompany impaired respiration in the cells of mutant staphylococci. The contents of catalase in the cells of mutants is low, and decarboxylase of diaminopimelic acid is apparently deleted. The latter deletion brings forth some alterations in the amino acid composition of the cell wall, where lysine is replaced by diaminopimelic acid.

A new pigment is observed in the cells of mutant staphylococci with maximal optical density at 475 mμ, which is absent in the parents. This points to the possibility that respiratory defects in mutant staphylococci are accompanied by some disturbances in the metabolism of flavines. As far as synthesis of vitamin B_{12} may be related to biogenesis of flavines (Woolley, 1954), it is interesting to record a sharp increase in the contents of vitamin B_{12} in the cells of mutant staphylococci as compared with their parents (Gause, Kochetkova & Vladimirova, 1961).

It is difficult to suggest at present which biochemical alterations in mutant staphylococci are of primary significance, and which might represent only the consequences of these primary alterations. In this connection

the available indirect evidence about differences related to deoxyribonucleic acid in parent and mutant staphylococci is of interest, as far as deoxyribonucleic acid is an important determinant of hereditary characters, including impaired respiration. It was observed that the growth of staphylococci with respiration deficiency is about 60 times more vulnerable to the inhibitory action of 5-fluorouracil than the growth of the initial parent culture, and in both cases this action is largely reversed by uracil, but not by thymine. On the other hand, 5-fluoro-2-deoxyuridine inhibits the growth of the parent culture of staphylococci much more strongly than does 5-fluorouracil, and this inhibition is reversed by thymine, but not by uracil. It is significant that 5-fluoro-2-deoxyuridine does not inhibit at all the growth of mutant staphylococci with respiration deficiency. In accordance with current views, 5-fluoro-2-deoxyuridine blocks the formation of thymine which is required for the synthesis of deoxyribonucleic acid (Cohen, Flaks, Barner, Loeb & Lichtenstein, 1958), and this points to the possibility of synthesis of deoxyribonucleic acid thymine in mutants through some pathway different from that of the parent strain (Gause, Kochetkova & Vladimirova, 1961). Another possible explanation for this phenomenon is that it may reflect the relative difficulty encountered by 5-fluoro-2-deoxyuridine in permeating to the site of its action in the mutant cell.

Biological Antimetabolites

Progression from metabolite action to antimetabolite potency in passing from one biochemical system to another similar but not identical biochemical system has been discussed by Woolley (1952) in great detail. Woods (1953) also reviewed a number of cases of inhibition of growth of cells by analogues of growth factors which are themselves metabolites in some other systems. This is due to the fact that essential metabolites are very exacting in their structural requirements. Even closely related molecules interfere with the normal functioning of the metabolite and are thus antimetabolites.

In view of these principles of antimetabolite inhibition, the components of biochemical systems of "analogues" of cancer cells in micro-organisms may in some cases behave as antimetabolites in the similar but not identical biochemical systems of cancer cells, as is shown on the scheme in the upper part of Fig. 1. In this respect the micro-organisms appear particularly promising, since the biochemical potentialities of microbes cover a wider spectrum than the cells of animal tissues.

Analysing this situation in greater detail one can say that in reality one deals with various kinds of tumours (A_1, A_2, A_3 . . . etc.), and with "cancer-like" mutants of micro-organisms belonging to different species (B_1, B_2,

B_3 . . . etc.). Taking into account the principles of antimetabolite inhibition one can expect, for example, that system B_2 may produce antimetabolite action in respect of systems B_1 and A_1 (Fig. 1 (b)). This would mean that a "cancer-like" mutant of micro-organisms B_2 produces antimetabolite which can be detected by its inhibitory action upon a "cancer-like" mutant micro-organism of another species, B_1. With the aid of such an assay the antimetabolite can be isolated and purified, and afterwards tested for inhibitory action upon a number of tumours in animals and man. One could expect that, for example, the isolated substance would reveal inhibitory action, i.e. behave as an antimetabolite, in respect of tumours of some definite type (say A_1).

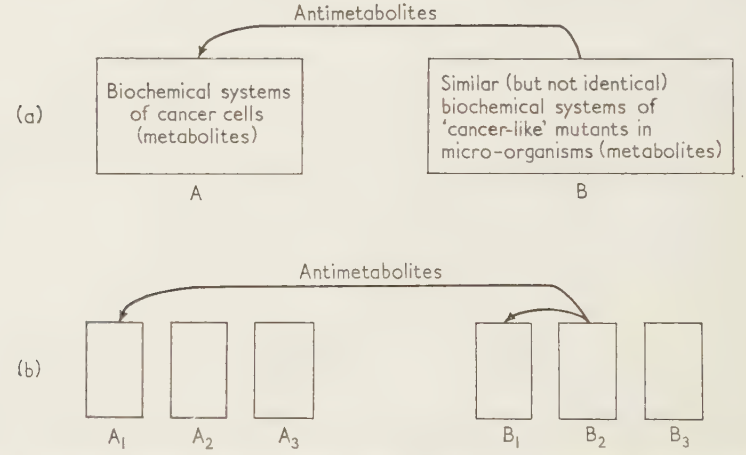


FIG. 1. Scheme showing the biological way of designing antimetabolites for malignant growth.

(*a*) General principle of designing.

(*b*) Various kinds of tumours (A_1, A_2, A_3 . . . etc.), and the "cancer-like" mutants in micro-organisms belonging to different species (B_1, B_2, B_3 . . . etc.).

The theoretic potentialities discussed above can be used for the development of new biological approaches in the search for antimetabolites of malignant growth. In mutant staphylococci with impaired respiration (mutant uv-2), described in detail by Gause (1960), an antimetabolite was observed which selectively inhibited the growth of mutant *Bacterium paracoli* with oxidation deficiency (mutant 43).

The antimetabolite produced by mutant staphylococci can only be recognized by using a test organism which is a respiration-deficient mutant belonging to another bacterial genus; the metabolite was found to be inactive when tested upon all available strains of bacteria possessing normal respiration (staphylococci, *B. coli*, *B. paracoli*, *B. subtilis*, *B. mycoides*, *Sarcina*, *Caryophanum*, and others). In other words, taking into

account designations used on the scheme in Fig. 1, one observes production of an antimetabolite by B_2, which selectively inhibits the growth of B_1.

Employing mutant B_1 as a test-organism (i.e. respiration-deficient culture of *B. paracoli* 43), the antimetabolite produced by mutant B_2 (i.e. staphylococci with impaired respiration belonging to the strain uv-2) was concentrated and purified, and its action upon several kinds of mouse tumours was studied. It has been recorded that this antimetabolite is not toxic, but that in mice with lymphoid leukemia it does not inhibit the growth of neoplasia. In the case of Crocker sarcoma of mice, the tumours were slightly inhibited by this substance, and their weight attained 75 to 80% of controls. In other words, the antimetabolite produced by mutant staphylococci and inhibiting mutants of *B. paracoli* with respiration deficiency is of no practical interest at present. But it appears that further studies along these lines could contribute to the development of new biological antimetabolites for malignant growth.

REFERENCES

Bennett, L. L., Skipper, H. E., Simpson, L., Wheeler, G. P. & Wilcox, W. S. (1960). *Cancer Res.* **20,** 62.

Cohen, S., Flaks, J., Barner, H., Loeb, M. & Lichtenstein, J. (1958). *Proc. nat. Acad. Sci., Wash.* **44,** 1004.

Danielli, J. F. (1959). "Squibb Centennial Lectures", p. 167. Putnam, New York.

Gause, G. F. (1959). Biol. Rev. **34,** 378.

Gause, G. F. (1960). "The Search for New Antibiotics". Oxford University Press, London; Yale University Press, New Haven.

Gause, G. F., Ivanitskaia, L. P. & Vladimirova, G. B. (1958). *Bull. Acad. Sci. U.R.S.S. Biol. Ser.* **6,** 719.

Gause, G. F. & Kochetkova, G. V. (1960). *Antibiotiki* (*Moscow*), **5,** 62.

Gause, G. F., Kochetkova, G. V. & Vladimirova, G. B. (1961). *C. R. Acad. Sci. U.R.S.S.* **139,** 223.

Haddow, A., de Lamirande, G., Bergel, F., Bray, R. & Gilbert, D. (1958). *Nature, Lond.* **182,** 1144.

Knuniants, J. L., Golubeva, N. E. & Kildisheva, O. V. (1960). *Adv. mod. Biol., Moscow,* **50,** 167.

Larionov, L. F. (1959). *Vestn. Akad. med. Nauk. S.S.S.R.* **6,** 25.

Monier, R., Zajdela, F., Chaix, P. & Petit, J. F. (1959). *Cancer Res.* **19,** 927.

Potter, V. R. (1958). Fed. Proc. **17,** 691.

Warburg, O. (1930). "The Metabolism of Tumours". Constable, London.

Warburg, O. (1956). *Science,* **123,** 309.

Weiler, E. (1959). "Ciba Foundation Symposium on Carcinogenesis: Mechanisms of action", p. 165. Churchill, London.

Weinhouse, S. (1956). *Cancer Res.* **16,** 654.

Woods, D. D. (1953). "Symposium on Nutrition and Growth Factors", p. 3. Istituto Superiore di Sanità, Roma.

Woolley, D. W. (1952). "A Study of Antimetabolites". Wiley and Sons, New York.

Woolley, D. W. (1953). *Proc. nat. Acad. Sci., Wash.* **39,** 6.

Woolley, D. W. (1954). "Symposium on Antimetabolites and Cancer", p. 163. Amer. Ass. Adv. Sci., Washington.

Zilber, L. A. (1960). *Adv. mod. Biol., Moscow,* **49,** 37.

Zilber, L. A., Abelev, G. J., Avenirova, S. A., Engelhardt, N. V. & Baidakova, S. L. (1959). *C. R. Acad. Sci. U.R.S.S.* **124,** 936.

J. Theoret. Biol. (1961) 1, 542–550

The Theory of Uncatalysed Linear Expanding Systems

C. J. Perret and H. C. Levey

Department of Microbiology and Department of Mathematics, University of Western Australia, Perth, Western Australia.

(*Received* 23 *March* 1961)

The kinetic theory of *expanding systems* is placed on firmer foundations by proving, for an extremely simple and conventionalized example, certain fundamental contentions previously supported only by non-mathematical arguments. Equations are developed that describe the conditions under which the uncatalysed expanding system will grow, and the manner in which concentration of "nutrient" will affect both growth-rate and the concentrations of "metabolites" before and during the time-independent *exponential state*.

The exponential and steady states are compared, and some biological implications of expanding-system theory are briefly discussed.

1. Introduction

Many aspects of the nature of biological growth are well illustrated by the properties of bacterial cultures, which provide simple, convenient and tractable examples of growing cell populations. For our present purpose the entire asynchronously dividing population within such a culture may be treated initially as a single system—the *biophase*—having a virtually constant surface-area–volume ratio maintained by the repeated division of its sub-units.

In a suitable environment a bacterial biophase continuously takes in nutrients from the surrounding medium, transforms them through a complex reticular sequence of metabolic steps, excretes some of the products back into the medium, and retains the remainder within the system. The inflow of material therefore exceeds the outflow; and the biophase undergoes *balanced growth*, during which both its mass and volume increase while the concentrations of metabolites remain relatively unchanged. In a constant environment the bacterial biophase grows exponentially with a specific growth rate usually proportional to low concentrations of nutrient, but tending towards a limiting maximal value as nutrient concentration is increased.

It has been customary to interpret the kinetics of unit volume of a growing biophase in terms of the steady-state of a constant-volume open system, in spite of the fact that, when the rate of growth of the biophase

is significant, some of the properties of the two systems are different. In addition, a capacity for balanced growth has been commonly regarded as peculiar to living systems; and previous formulations of the phenomenon in physico-chemical terms (by either the "autosynthetic integration" or "self-replicating molecule" approaches) seem to imply that such systems must, of necessity, possess a degree of complexity not substantially less than that found in known unicellular micro-organisms. However, it has recently been suggested that balanced growth can be interpreted in terms of the simple reaction kinetics of *expanding systems*, and that the properties of such systems are strikingly similar to the observed properties of a growing biophase (Perret, 1959, 1960).

The crucial assumption of the expanding system concept is that the volume of an open system, instead of being fixed, could be proportional to the amount of one or more retained intermediates; there would then be a potentially infinite *internal sink* for all components of the system. The significance of the properties of expanding systems and some of the ways in which they may assist our understanding of the kinetics of growth and the metabolism of growing cells, and possibly of the nature of biopoesis or biogenesis, have already been discussed at length (Perret, 1959, 1960). In this paper we attempt to place the theory of expanding systems on firmer foundations by proving, for an extremely simplified and conventionalized example, some of the fundamental contentions previously supported only by non-mathematical arguments (Perret, 1960).

2. Theory

ATTRIBUTES OF THE SIMPLE MODEL SYSTEM

An expanding system and the medium surrounding it can be miscible liquid phases if the volume-controlling components of the system function by forming a semi-permeable boundary membrane. But in the example considered here (Fig. 1) we suppose, for the sake of simplicity, that system and medium are two immiscible liquid phases each "perfectly mixed" within itself at all times, so that there are no concentration gradients except at the interface. The conventions of Fig. 1 are those generally used for chemical open systems, except that the boundaries are indicated by

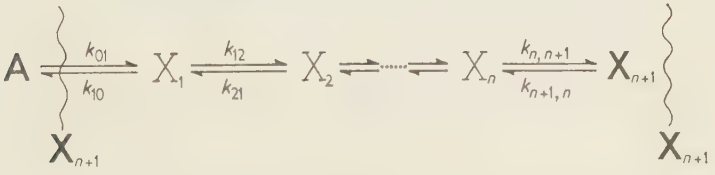


FIG. 1.

wavy vertical lines terminated by the symbol of the volume-controlling component, and the "source" and "sink" materials are represented by bold face capitals (Perret, 1960).

The surface area of the system is assumed to be proportional to the volume V; and both are proportional to the total *amount* of component (metabolite) $\mathbf{X}_{n+1}$, which forms the solvent in which all the other components are dissolved while being itself negligibly diluted by the solutes; its concentration therefore does not change. The medium serves as an *infinite* reservoir of the source material (nutrient) **A**, which is in solution at concentration a.

Metabolite $\mathbf{X}_{n+1}$ is formed from nutrient **A** by a linear sequence of first-order reactions, but there is no *external* sink, and the $\mathbf{A} \leftrightarrow X_1$ transformation is the only means by which material can pass through the interface; thus all other metabolites and $\mathbf{X}_{n+1}$ itself are retained within the system. In physical terms this particular example of an expanding system could be pictured as a phase with a constant surface-area–volume ratio into which **A** is entering from the medium by diffusion; X_1 would then be *internal* **A**, which is changing spontaneously into a series of compounds soluble in the system phase but negligibly soluble in the medium.

DERIVATION OF THE RATE EQUATIONS

We may now proceed to write down the rate equations for each metabolite. However, the usual rate equations (see, e.g., Kacser, 1957) for the kinetics of homogeneous reactions apply to constant-volume systems, in which concentrations of components are altered only by processes of formation or decomposition; such equations therefore apply equally to changes in both the *concentration* and the *amount* of a reactant. But in a system whose volume is variable, concentrations can change while amounts remain constant; thus if we denote by X_i and x_i $(= X_i/V)$ respectively the amounts and concentrations, we have for reactant X_i $(i = 1, \ldots, n+1)$ the additional "sink-like" term

$$\lim_{\delta t \to 0} \frac{1}{\delta t}\left(\frac{X_i}{V+\delta V} - \frac{X_i}{V}\right) = -\frac{x_i}{V}\frac{dV}{dt} \qquad (1)$$

in the equation for the rate of change of its concentration.

This, of course, is equivalent to the statement that whether or not the volume is constant, the usual rate equation applies to the *amounts* of reacting substances. If we now denote by $k_{i,\,i+1}$, $k_{i+1,\,i}$ $(i = 0, \ldots, n)$ the specific rate constants for the forward and reverse transformations, a typical rate equation for *concentrations* is in its more general form,

$$\frac{dx_i}{dt} = k_{i-1,\,i}x_{i-1} - (k_{i,\,i-1} + k_{i,\,i+1})x_i + k_{i+1,\,i}x_{i+1} - \frac{x_i}{V}\frac{dV}{dt} \qquad (2)$$

(the last term vanishing for constant-volume systems) which is the same as the rate equation for *amounts*

$$\frac{dX_i}{dt} = k_{i-1,\,i}X_{i-1} - (k_{i,\,i-1} + k_{i,\,i+1})X_i + k_{i+1,\,i}X_{i+1} \tag{2a}$$

Since the system has a constant surface-area–volume ratio the *amount* of nutrient **A** entering the system in unit time is proportional to aV, and

$$V = \rho^{-1} X_{n+1} \tag{3}$$

where ρ is a constant.
Hence the complete set of rate equations for *amounts* in the system in Fig. 1 is

$$\frac{dX_1}{dt} = -(k_{10} + k_{12})X_1 + k_{21}X_2 + k_{01}a\rho^{-1}X_{n+1} \tag{4}$$

$$\frac{dX_i}{dt} = k_{i-1,\,i}X_{i-1} - (k_{i,\,i-1} + k_{i,\,i+1})X_i + k_{i+1,\,i}X_{i+1} \quad i = 2, \ldots, n \tag{5}$$

$$\frac{dX_{n+1}}{dt} = k_{n,\,n+1}X_n - k_{n+1,\,n}X_{n+1} \tag{6}$$

The system is stationary, in a state of chemical equilibrium, when the rate of change of *amount* of every metabolite is zero. We now assume that the system has been set up at equilibrium with a reservoir of **A** at concentration a', and that the values of the X_i are X_i'. The following set of relations is then easily derived:

$$k_{i-1,\,i}X'_{i-1} = k_{i,\,i-1}X'_i \qquad i = 2, \ldots, n+1 \tag{7}$$

and

$$k_{01}X'_{n+1}a'\rho^{-1} = k_{10}X'_1 \tag{8}$$

In order to investigate the manner in which such a system will respond to changes in the concentration of **A** it is convenient to scale the variables X_i and a according to their values in the initial stationary state. Thus, if

$$\xi_i = X_i/X'_i$$
$$\alpha = (a - a')/a' \qquad (\geqslant -1) \tag{9}$$

then for the system at equilibrium, α, the *effective concentration* of nutrient is zero; and ξ_i, the *effective amounts* of the metabolites are all unity. The rate equations now take the form

$$\frac{d\xi_1}{dt} = -(\mu_0 + \lambda_0)\xi_1 + \lambda_0\xi_2 + (1 + \alpha)\mu_0\xi_{n+1} \tag{10}$$

$$\frac{d\xi_i}{dt} = \mu_{i-1}\xi_{i-1} - (\mu_{i-1} + \lambda_{i-1})\xi_i + \lambda_{i-1}\xi_{i+1} \qquad i = 2, \ldots, n+1$$

where

$$\begin{aligned} \mu_{i-1} &= k_{i,\,i-1} \geqslant 0 \qquad i = 1, \ldots, n+1 \\ \lambda_{i-1} &= k_{i,\,i+1} \geqslant 0 \qquad i = 1, \ldots, n \\ \lambda_n &= 0 \end{aligned} \tag{11}$$

This system may be written in the matrix form

$$\frac{d\xi}{dt} = A(\alpha)\xi \tag{12}$$

where

$$A(\alpha) = \begin{pmatrix} -(\mu_0 + \lambda_0) & \lambda_0 & . & \ldots & (1+\alpha)\mu_0 \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & \ldots & . \\ . & & & & \\ . & & & & \\ . & & & \mu_n & -\mu_n \end{pmatrix} \tag{13}$$

CONDITIONS FOR GROWTH OF THE SYSTEM

Our aim now is to show that if α is a positive (negative) constant the system will grow (shrink); and clearly, by a continuity argument, this depends in general on whether the stationary state ($\alpha = 0$) is stable to small disturbances in the ξ_i from their stationary values, which are unity. All of these questions are decided by the latent roots of $A(\alpha)$; for, by use of either the Laplace transform or the method of normal modes, the solution is, in general, a linear combination of terms $e^{p_j t}$, where p_j is the jth latent root of $A(\alpha)$.

The latent roots of $A(\alpha)$ are the roots of the characteristic polynomial

$$\begin{aligned} P(p, \alpha) &= det(pI - A(\alpha)) \\ &= p^{n+1} + A_1 p^n + \ldots + A_{n+1} \end{aligned} \tag{14}$$

where $A_{n+1} = det(-A(\alpha))$. Now when $det(pI - A(\alpha))$ is expanded from the first row it follows that α appears only in the constant term of the polynomial, namely A_{n+1}, while the terms independent of α are those from the continuant obtained by deleting the top right-hand term $-(1+\alpha)\mu_0$. The contribution from this top right-hand term is easily seen to be

$$\begin{aligned} &-(1+\alpha)\mu_0(-1)^{n+2}.(-1)^n \mu_1\mu_2 \ldots \mu_n \\ &= -(1+\alpha)\mu_0\mu_1 \ldots \mu_n \end{aligned} \tag{14a}$$

The continuant is most easily dealt with by means of a recurrence relationship established as follows. We denote by $S_{n+1-r}(p)$ the determinant

$$\begin{vmatrix} \mu_r+\lambda_r+p & -\lambda_r & \cdot & \cdots & \cdot \\ -\mu_{r-1} & \mu_{r-1}+\lambda_{r-1}+p & -\lambda_{r+1} & \cdots & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & -\mu_n & \mu_n+p \end{vmatrix} \tag{15}$$

and then by expansion from the first column

$$S_{n+1-r}(p)=(\mu_r+\lambda_r+p)S_{n-r}(p)-\lambda_r\mu_{r+1}S_{n-r-1}(p) \qquad r=0, \ldots, n-2 \tag{16}$$

where
$$S_1(p)=\mu_n+p \tag{17}$$

In this notation, the characteristic polynomial is given by

$$P(p, \alpha)=S_{n+1}(p)-(1+\alpha)\mu_0\mu_1 \cdots \mu_n \tag{18}$$

It is easily established by induction that

$$S_r(0)=\mu_{n+1-r} \cdots \mu_n$$

and hence
$$S_{n+1}(0)=\mu_0\mu_1 \cdots \mu_n \tag{19}$$

Also by use of the technique of Ledermann and Reuter (1954) it may be shown that the r roots of the polynomial $S_r(p)$ are real, distinct and negative. In particular, the $(n-1)$ roots of $S_{n+1}(p)$ are real, distinct and negative, and so the n roots of $S'_{n+1}(p)$ are also real, distinct and negative. If the algebraically greatest root of $S'_{n+1}(p)$ is $-\epsilon$, then $S_{n+1}(p)$ is monotonic increasing for $p>-\epsilon$ and positive for $p \geqslant 0$. Hence also $P(p, \alpha)$ is monotonic for $p>-\epsilon$ and since, from (14a) and (19),

$$P(0, \alpha)=-\alpha\mu_0\mu_1 \cdots \mu_n \tag{20}$$

(verifying the zero latent root for $\alpha=0$), then for $\alpha>0$ there is certainly one positive real root. Thus the system will grow when α is positive. Likewise, for small enough negative α the system will shrink.

For further information about the growth rate, and the stability of the stationary state, more needs to be known about the roots of $P(p, \alpha)$. Unfortunately, in contradistinction to the roots of $S_{n+1}(p, \alpha)$, these roots need not be real, as may be shown by simple examples; thus the one positive real root obtained above may not be the only root with positive real part. However, for small n, it is possible to investigate the distribution of the roots of $P(p, \alpha)$ directly by a criterion of the Hurwitz type (Marden, 1949) and it may be shown that for $n \leqslant 3$ there is only one real root with positive real part when $\alpha>0$; while for $-1 \leqslant \alpha<0$ all of the roots have negative real parts. When $\alpha=0$ the zero root is the root with greatest real part so the equilibrium state is neutrally stable. Values of $\alpha<-1$ have no physical meaning.

On the basis of the above, we postulate that for all n this system is stable in the stationary state and that when α is positive it has only one latent root with positive real part. It follows that this root is real, and that it is the asymptotic growth rate, $\bar{\gamma}$. For large α it is given by

$$\bar{\gamma} = p_{n+1} = \alpha^{1/n+1}(\mu_0\mu_1 \ldots \mu_n)^{1/n+1}[1 + O(\alpha^{-1/n+1})] \tag{21}$$

and for small α

$$\bar{\gamma} = p_{n+1} = \alpha(\mu_0\mu_1 \ldots \mu_n/A_n)[1 + O(\alpha)] \tag{22}$$

(The properties of the roots of $S_{n-1}(p)$ imply $A_1, \ldots, A_n > 0$). It also follows from the strict monotony of $P(p, \alpha)$ for $p > 0$ that $\bar{\gamma}$ increases monotonically with α.

EFFECT OF NUTRIENT CONCENTRATION ON THE RELATIVE CONCENTRATIONS OF METABOLITES

Finally, we show that in such a growing system the asymptotic ratio of the amounts (and therefore of the concentrations) of the rth to the $(r+1)$th metabolite increases as α increases. For since there is only one positive latent root $\bar{\gamma}$ then ultimately

$$\xi_i \sim c_i e^{\bar{\gamma}t} \qquad i = 1, \ldots, (n+1)$$

where the c_i are constants determined by the initial conditions, and in fact form the elements of the latent vector corresponding to the root $\bar{\gamma}$. By substitution in the equations (10) it follows that

$$(\lambda_{i-1} + \mu_{i-1} + \bar{\gamma})c_i = \mu_{i-1}c_{i-1} + \lambda_{i-1}c_{i+1}, \qquad i = 2, \ldots, n+1$$

that is

$$\frac{c_{i-1}}{c_i} = \frac{\lambda_{i-1} + \mu_{i-1} + \bar{\gamma}}{\mu_{i-1}} - \frac{\lambda_{i-1}}{\mu_{i-1}}\frac{c_{i+1}}{c_i} \qquad i = 2, \ldots, n \tag{23}$$

Hence,

$$\frac{\partial}{\partial\bar{\gamma}}\left(\frac{c_{i-1}}{c_i}\right) = \frac{1}{\mu_{i-1}} - \frac{\lambda_{i-1}}{\mu_{i-1}}\frac{\partial}{\partial\bar{\gamma}}\left(\frac{c_{i-1}}{c_i}\right)$$

and if $\dfrac{\partial}{\partial\bar{\gamma}}\left(\dfrac{c_i}{c_{i+1}}\right) > 0$ then so is $\dfrac{\partial}{\partial\bar{\gamma}}\left(\dfrac{c_{i-1}}{c_i}\right)$. But from the equation with $i = n+1$

$$\frac{\partial}{\partial\bar{\gamma}}\left(\frac{c_n}{c_{n+1}}\right) = \frac{1}{\mu_n} > 0$$

therefore, since $\partial\bar{\gamma}/\partial\alpha > 0$, all of the asymptotic metabolite concentration gradients

$$\xi_i/\xi_{i+1} \qquad i = 1, \ldots, n$$

increase as α increases.

3. Conclusions and Discussion

The preceding calculations support the contention that an uncatalysed linear expanding system having a constant surface-area–volume ratio differs in respect of the following significant properties from a comparable constant-volume open system.

(i) The expanding system grows in mass and volume when the effective concentration of nutrient is positive (i.e. when the absolute concentration is above the level which would be at equilibrium with the concentration of the volume-controlling metabolite). In a constant environment the expanding system then tends asymptotically towards an *exponential state* in which both mass and volume increase indefinitely at an unchanging exponential rate. Under similar circumstances a constant-volume open system tends towards the steady-state condition of constant mass.

(ii) During the exponential state the rate of formation of every metabolite exceeds its rate of decomposition in such a manner that its *amount* increases exponentially while its *concentration* remains constant; whereas rates of formation and decomposition are equal in the steady-state, and amounts and concentrations are both constant.

(iii) Raising the effective nutrient concentration increases the ratio of the exponential-state concentrations of any pair of metabolites taken in order of increasing *kinetic distance* from the source; the proportional composition of the system therefore varies with the growth rate. In contrast, the proportional composition of an open system in the steady state is independent of the flux rate (Kacser, 1957).

(iv) The specific growth rate of the expanding system also increases as the effective concentration of nutrient is raised; but *equal* positive increments in nutrient concentration lead ultimately to *progressively diminishing* positive increments in growth rate. In the comparable constant-volume system the flux-rate during the steady state remains proportional to effective concentration of nutrient.

The resemblance between the behaviour of a growing bacterial biophase and the behaviour of an expanding system even as simple as that represented in Fig. 1 is therefore obvious. It appears that inclusion of an autocatalytic stage in the system could considerably increase its potentialities, and we are at present seeking a mathematical treatment of that modification.

Since the primary purpose of the expanding-system concept is to provide a kinetic "model" of a growing biophase there is no need for its simplest idealized formulations to have any physical counterpart. It is therefore permissible, as a first approximation, to ignore the formation of concentration gradients within and around the system; and to offer no explanation of the means by which the hypothetical model could overcome the *scale*

effect and maintain a constant surface-area–volume ratio. But the realization that systems lacking the interrelated molecular and structural complexities of known cells can theoretically exhibit balanced growth leads naturally to speculations about the possibility of their real existence.

The first step in the transition from theory to reality seems to be the assumption that the expanding system consists, not of a single continuous phase, but of a large population of metabolizing, growing and dividing droplets, each with its associated concentration gradients. Rashevsky (1938, 1960) has already cogently argued that the diffusion forces in metabolizing droplets could lead to their regular division; so a population of such droplets might maintain their individuality and increase in numbers, instead of tending to coalesce like the disperse phase of an "inanimate" emulsion. We certainly do not claim at present that an extended and modified theory of expanding systems could be combined with Rashevsky's relatively neglected ideas to form the basis of a general physico-chemical theory of life; nor would that theory, if developed, be other than one of a number of possible alternatives. Nevertheless, it is intriguing to realize that the three essential prerequisites for primitive evolution—balanced growth, division and heritable variation—may soon be explicable in terms of the kinetics of such familiar phenomena as chemical reaction, partition, diffusion and interfacial tension. Meanwhile the expanding system concept apparently has the advantage that, in addition to its primary theoretical significance, it indicates how some possible aspects of the origin of life might be experimentally investigated in artificial systems.

REFERENCES

KACSER, H. (1957). Appendix to WADDINGTON, C. H. "The Strategy of the Genes" George Allen & Unwin, London.

LEDERMANN, W. & REUTER, G. E. H. (1954). *Phil. Trans.* **A246,** 321.

MARDEN, M. (1949). "The Geometry of the Zeros of a Polynomial in a Complex Variable", Chapter 9. American Mathematical Society, New York.

PERRET, C. J. (1959). *Aust. sci. Teachers' J.* **5,** 9.

PERRET, C. J. (1960). *J. gen. Microbiol.* **22,** 589.

RASHEVSKY, N. (1938). "Mathematical Biophysics", 1st Edition. Chicago University Press, Chicago; (1960) *ibid.* 3rd Revised Edition. Dover Publications, New York.

Note of Acknowledgment

Evolution and the Theory of Games

Through an inexcusable oversight my paper on "Evolution and the Theory of Games", in the last issue of this *Journal*, failed to point out that it is to C. H. Waddington that we owe the original suggestion of the usefulness of game theory in evolutionary problems. The germinal suggestion of Waddington is contained in "The Strategy of the Genes" and indeed the whole distinction between strategies and tactics in evolution is a crucial point of that book. My debt to Waddington was acknowledged in an early sketch of the paper, but in some way it was omitted from the final manuscript.

It should also be pointed out that Dr Richard Levins has simultaneously and independently, through the use of convex set theory, attacked many evolutionary problems from essentially the same standpoint. I am vastly indebted to him for many fruitful discussions contrasting and comparing our two approaches.

R. C. Lewontin
Zoology Department
University of Sydney
Australia

Instructions to Authors

1. The *Journal of Theoretical Biology* is intended to be a central publishing medium for theoretical work in all fields of biology.

Appropriate subjects for papers will include:

(*a*) generalized theories;
(*b*) theories of specific processes or phenomena, e.g. crossing over, learning, active transport, photosynthesis, carcinogenesis;
(*c*) theoretical discussion of specific projects, e.g. selective mutation, drug design, electronic models;
(*d*) theoretical discussion of methods, e.g. cytochemical methods and instrumentation, amplification techniques;

Papers concerned mainly with experimental work will not normally be accepted, but a paper will not be rejected simply because it contains some new experimental material.

2. Reviews will be accepted only if they include a substantial new theoretical discussion. Letters to the editor, intended to direct attention to a particular theoretical point, will be accepted.

3. *Submission of Manuscripts.* Manuscripts should be sent to Professor J. F. Danielli, Zoology Department, King's College, London, or to any member of the Editorial Board. The submission of manuscript will be taken to imply that the material is original.

Papers may be submitted in English, French or German. It is emphasized however that papers in English will reach a much larger group of readers than those in other languages. A paper submitted in French or German should have an abstract in English.

Publication will be accelerated if two copies of a manuscript are submitted.

4. *Manuscripts.* If the Journal is to serve its intended purpose, it is essential that the main body of its contents should be intelligible to all biologists. To this end, wherever possible mathematical development should be confined to appendices, and specialized terms should be clearly defined in the introduction to a paper.

Summary. Every paper should be preceded by a summary: this should not normally exceed 300 words, and should be intelligible to the general reader without reference to the main text. Abbreviations should not be used.

Lay-out. Manuscripts should be typewritten in double spacing throughout (including summary, footnotes, tables and legends) on one side of quarto paper, with wide margins. Footnotes, tables and legends for illustrations should be typed separately at the end of the manuscript.

Running Headline. A short running title of not more than 40 characters (including spaces), suitable for page headings, should be given if the full title is longer than this.

5. *Conventions.* The conventions used in current issues of the Journal for headings, references, etc., should be followed in preparing manuscripts. References should be listed alphabetically at the end of the paper, and abbreviated according to the *World List of Scientific Periodicals.* Citations in the text should read thus: Smith, Brown & Robinson (1957), or (Smith, Brown & Robinson, 1957). When a citation has more than two authors, the names of all of them should be given at the first citation (except for more than five authors), and on subsequent citations thus:

(Smith *et al.*, 1957). The conventions (Brown, 1957a), (Brown, 1957b) should be used when more than one paper by the same author(s) has appeared in one year. Volume number should be indicated in the typescript thus: 53.

Footnotes should be avoided whenever possible. Customary abbreviations will be used; authors may refer to *Chemical Abstracts*, to the *Biochemical Journal*, **66**, 8 (1957), and to *Notes on the preparation of papers to be communicated to the Royal Society* (1951). Certain non-standard abbreviations of chemical substances are convenient, and those listed in the *Journal of Biological Chemistry* may be used without further definition (e.g. DNA, ATP). Any other abbreviations of this kind should be used sparingly, and must be defined in a footnote at the point in each manuscript where the first such abbreviation occurs.

6. *Illustrations.* It is often convenient to make a drawing two or three times larger than it will appear in the Journal, but very large drawings should be avoided and it will be helpful if drawings and illustrations are submitted on quarto sheets. Drawings should be made in Indian ink on tracing linen, Bristol board, or faintly blue-lined graph paper; the originals should always be submitted with the manuscript, since photo-prints of drawings cannot be reproduced satisfactorily.

Line drawings will be referred to as Fig. 1, Fig. 2, etc., and photographs as Plate I, Plate II, etc. The name(s) of the author(s) and the number should be written on all illustrations, together with the indication "Top" if necessary. Besides the original illustrations *one* duplicate set must accompany the second copy of the manuscript; these may be photo-prints, and need not be of the same quality as the originals but must be clear and legible. If oversize original drawings are submitted (bigger than 8 by 10 inches) they must be accompanied by *two* sets of small photo-prints.

7. *Reprints.* Fifty reprints of each paper will be supplied free of charge; additional copies may be purchased. Order forms will be sent with proofs.

Index to Volume 1

Journal of Molecular Biology

Published at 128s. 6d. ($18.00) per annual volume of 6 issues

Volume III, Number 5, October 1961

CONTENTS

ACADEMIC PRESS INC.
LONDON NEW YORK

Contents

Printed in Great Britain by The Whitefriars Press Ltd., London and Tonbridge